

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

March 6, 2001, 11:50:37 ; Search time 47.68 Seconds

Title:	US-09-407-430-1
Perfect score:	001

Sequence: 1 MNSKGQYPTQPTYPVQPPGN.....VTORKGNFEMGGSDGGYTIW 168

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

```

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

```

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 08
Maximum Match 100%

Listing first 45 summaries

Database :

121	/SIDS1/gcgdata/genseq/genseqp/A11980.DAT.*
122	/SIDS1/gcgdata/genseq/genseqp/A11981.DAT.*
123	/SIDS1/gcgdata/genseq/genseqp/A11982.DAT.*
124	/SIDS1/gcgdata/genseq/genseqp/A11983.DAT.*
125	/SIDS1/gcgdata/genseq/genseqp/A11984.DAT.*
126	/SIDS1/gcgdata/genseq/genseqp/A11985.DAT.*
127	/SIDS1/gcgdata/genseq/genseqp/A11986.DAT.*
128	/SIDS1/gcgdata/genseq/genseqp/A11987.DAT.*
129	/SIDS1/gcgdata/genseq/genseqp/A11988.DAT.*
130	/SIDS1/gcgdata/genseq/genseqp/A11989.DAT.*
131	/SIDS1/gcgdata/genseq/genseqp/A11990.DAT.*
132	/SIDS1/gcgdata/genseq/genseqp/A11991.DAT.*
133	/SIDS1/gcgdata/genseq/genseqp/A11992.DAT.*
134	/SIDS1/gcgdata/genseq/genseqp/A11993.DAT.*
135	/SIDS1/gcgdata/genseq/genseqp/A11994.DAT.*
136	/SIDS1/gcgdata/genseq/genseqp/A11995.DAT.*
137	/SIDS1/gcgdata/genseq/genseqp/A11996.DAT.*
138	/SIDS1/gcgdata/genseq/genseqp/A11997.DAT.*
139	/SIDS1/gcgdata/genseq/genseqp/A11998.DAT.*
140	/SIDS1/gcgdata/genseq/genseqp/A11999.DAT.*
141	/SIDS1/gcgdata/genseq/genseqp/A12000.DAT.*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	127	13.8	106	21	Y65155	Human 5' EST rela
2	116.5	12.6	1274	20	M89253	Human ALP. Homo
3	115.5	12.5	264	19	W71219	Beta-D-galactosid
4	114.5	12.4	258	19	M61954	Rat galectin amin
5	111.5	12.1	572	18	M31855	Mycobacterium tub
6	111.5	12.1	763	18	M31852	Mycobacterium tub
7	110	11.9	214	17	R86913	Cotton fiber-spec
8	108.5	11.8	259	10	P91655	Elmeria cell surf
9	108	11.7	262	14	RA42200	IgE binding prote
10	108	11.7	725	20	Y59724	Human normal ova
11	108	11.7	181	21	Y82707	Human apoptos
12	108	11.7	1220	21	Y82708	Human apoptos

13	107.5	11.7	264	12	R1325
14	107.5	11.4	278	12	R1333
15	105	11.4	114	13	R2916
16	105	11.4	259	17	R9769
17	105	11.4	1185	20	Y3349
18	104	11.3	302	20	Y0261
19	103.5	11.2	751	16	R0803
20	103.5	11.2	902	17	Y0225
21	102.5	11.1	203	9	P8290
22	102.5	11.1	203	9	P83193
23	102.5	11.1	1291	20	Y0168
24	102.5	11.1	1291	20	Y0681
25	102	11.1	174	19	M81722
26	102	11.1	174	19	M64366
27	102	11.1	174	20	Y39155
28	102	11.1	174	20	Y39015
29	102	11.1	492	20	Y27244
30	102	11.1	626	18	W15399
31	101.5	11.0	902	20	M85730
32	101.5	11.0	1248	13	Y14666
33	101	11.0	171	13	R27476
34	101	11.0	302	20	Y02614
35	101	11.0	302	20	Y02612
36	101	11.0	302	20	Y08458
37	101	11.0	302	20	Y08458
38	100.5	10.9	302	20	Y08451
39	100.5	10.9	2441	16	R79054
40	100	10.9	2441	21	Y94222
41	100	10.9	303	18	Y22448
42	100	10.9	303	18	M23280
43	100	10.9	303	18	M18682
44	100	10.9	303	20	M64320
45	100	10.9	303	20	Y39122
					Y18965

ALIGNMENTS

RESULT	1
ID	Y65155 standard; Protein; 106 AA.
XX	Y65155
AC	Y65155;
XX	
DT	01-FEB-2000 (first entry)
XX	
DE	Human 5' EST related polypeptide SEQ ID NO:1316.
XX	
KW	Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;
KM	gene therapy; chromosome mapping; upstream regulatory sequence;
KW	forensic; location; development; protein synthesis; stability;
KM	regulation; Identification.
XX	
OS	Homo sapiens.
XX	
PN	MO9953051-AZ.
XX	
PD	21-OCT-1999.
XX	
FE	09-APR-1999; 99WO-IB00712.
XX	
PR	09-APR-1998; 98US-0057719.
PR	28-APR-1998; 98US-0065047.
XX	
PA	(GEST) GENSET.
PI	Dumas Milne Edwards J, Duclert A, Giordano J;
DR	N-PDB; Z42769.
PT	Novel secreted protein 5' expressed sequence tag sequences used in diagnostic, forensic, gene therapy, and chromosome mapping procedures.

XX New nucleic acid encoding specific protein tyrosine phosphatases
PT useful for identifying specific modulators for treatment and
PT prevention of cancer and neurodegenerative disease
PS
PS Claim 2; page 160-164; 193pp; English.
XX
XX The present invention describes isolated, enriched or purified nucleic
CC acids encoding PTP04, SMD, PTP05, PTP10, ALP and ALK-7 proteins. The
CC present sequence represents human ALP. The above proteins, other than
CC ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify
CC substances that modulate their activity (i.e. agonists and antagonists,
CC including NMP) in vivo or in vitro. These substances are used to treat
CC or prevent diseases associated with abnormal signal transduction
CC pathways that involve the proteins, particularly cancer (e.g. leukemia
CC and lymphoma), while modulators of ALK-7 (which is a type I receptor
CC serine/threonine kinase) are used to promote neuronal survival,
CC particularly for treating Alzheimer's, Parkinson's or Huntington's
CC diseases. Nucleic acid fragments of the polynucleotides encoding the
CC proteins can be used as probes to identify and clone related sequences
CC to detect protein-encoded RNA; to generate transgenic animals and in
CC gene therapy (optionally after mutation). Ab are used to determine the
CC proteins.
XX
SQ Sequence 1274 AA.

Query Match	12.68;	Score 116.5;	DB 20;	Length 1274;
Best local Similarity	29.44;	Pred. No. 0.033;		
Matches 47;	Conservative 12;	Mismatches 54;	Indels 47;	Gaps 9;

Oy 7 YPTQPIRYVNPQPGNVPYQTLLH-----POAP-----PYDA-----PPASESLYR 47
 :::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 600 hpsqgaigr-qppqgqlpqhnhlfpqpaqvllppspvyvapepgvvlgqprrplntfily- 65

104 VLVGGYDAGARF-----GAGATAGNIPPPPPGCPP 134

W71219

ID	W11219	Standard; Floetm, 207 mm.
XX		
AC	W71219;	
XX		
DT	30-OCT-1998	(first entry)

DE Beta-D-galactoside-binding protein designated L-31-gal-lectin

KW Beta-D-galactoside-binding protein; L-3I-gal-lectin
antibody

XX:

XX 00
XXXX

PN 055801002-A.
XY

01-SEP-1998. PD

PF 22-NOV-1995;

PR 26-JAN-1994;

PR 06-JAN-1989;
PR 05-APR-1991;

PR 22-NOV-1995; 9505-0306311.

AA
PA (KARM-) KARMANOS CANCER INST BARBARA ANN

Part A.

XX

DT

11

XX WPI: 1998-494766/42.
 DR N-PSDB: V54736.
 XX
 PT Test for metastatic potential of cell sample - by measuring binding
 of antibody to L-31-gal-lectin on cell surface
 XX
 PS Disclosure: Fig 3A-B; 24pp; English.
 CC The present sequence represents a beta-D-galactoside-binding protein
 CC designated L-31-gal-lectin. The sequence is derived from clone 2. The
 CC specification describes a method for testing a cell sample for
 CC metastatic potential. The method comprises contacting the sample with
 CC a labelled antibody that binds to endogenous cell-surface
 CC L-31-gal-lectin, removing unbound antibody, and determining the amount
 CC of bound antibody as a measure of L-31-gal-lectin expression, where the
 CC metastatic potential increases as the level of L-31-gal-lectin
 CC expression increases. The antibody is produced by immunisation with a
 CC L-31-gal-lectin protein.
 XX
 SQ Sequence 264 AA:

Query Match 12.5%; Score 115.5; DB 19; Length 264;
 Best Local Similarity 32.5%; Pred. No. 0.0074;
 Matches 54; Conservative 8; Mismatches 49; Indels 55; Gaps 13;

OY 5 GQPTQP---TYPVO-PPGNPVYPTLHLPOAPRYTDAPRAYSELYRSPVHPCATVPT 60
 DB 42 ggyppqgqppgyppgqppg---gyp-----gqap-----psay-----pgrtapgqaypgrt 84
 OY 61 MSAAFGASLYLPMASVAVGRLGSTITPMAYYPVGPYTP--PGSTVLVEGCGYDAGARFGA 118
 DB 85 apgqayp-----gslapqaf-----pgqpgqgqaypypagqayp 123
 OY 119 GATAGNIP---PPPGCPNMAQLAVMKG-----ANLVYTO-RKGN 155
 DB 124 psgpltpydpdpdgglmptr--mltlmglvkpnantivldfrrgn 167

RESULT 4
 ID W61954 standard; protein; 258 AA.
 XX
 AC W61954;
 XX
 DT 18-SEP-1998 (first entry)
 XX
 DE Rat galectin amino acid sequence.
 XX
 KM Mortalin; galectin; diabetes-mediating protein; insulin; DMP;
 XX diabetes; drug screening assay.
 OS
 XX Rattus sp.
 PN W09820124-A2.
 XX
 PD 14-MAY-1998.
 XX
 PF 24-OCT-1997; 97WO-IB01627.
 XX
 PR 18-JUL-1997; 97US-0897098.
 PR 25-OCT-1996; 96US-0029324.
 PR 05-NOV-1996; 96US-0030088.
 PR 05-NOV-1996; 96US-0030186.
 XX
 PA (ANDE/) ANDERSEN H U.
 PA (CHRI/) BJERRE CHRISTENSEN U.
 PA (FEYS/) FEY S J.
 PA (KARL/) KARLSEN A E.
 PA (LARS/) MOSE LARSEN P.
 PA (NERU/) NERUP J.
 PA (POCI/) POCTOT F.

XX
 PI Andersen HU, Bjerre Christensen U, Fey SJ, Karlson AE;
 PI MOSE LARSEN P, Nerup J, Poctot F;
 XX
 DR WPI: 1998-286940/25.
 PT Identification of diabetes-mediating protein(s) - by transplanting
 PT insulin-secreting cells into host at risk of developing diabetes and
 PT analysing protein expression in transplanted cells
 XX
 PS Disclosure: Fig 4; 154pp; English.
 CC This represents the amino acid sequence of murine mortalin. This is a
 CC diabetes-mediating protective protein used in the method of invention.
 CC The invention provides methods for in vivo identification of a diabetes-
 CC mediating protein (DMP) by transplanting insulin-secreting cells into
 CC host at risk of developing diabetes and analysing protein expression in
 CC transplanted cells. The DMPs are useful in drug screening assays for
 CC identifying compounds capable of modulating the development of diabetes,
 CC useful as therapeutic agents for the treatment or prevention of diabetes,
 CC and useful as targets of therapeutic agents capable of preventing or
 CC ameliorating diabetes by modulating the expression of the DMP. Changes in
 CC the expression of specific DMPs is diagnostically useful as indicator of
 CC the development of diabetes.
 XX
 SQ Sequence 258 AA:

Query Match 12.4%; Score 114.5; DB 19; Length 258;
 Best Local Similarity 31.5%; Pred. No. 0.0088;
 Matches 51; Conservative 5; Mismatches 63; Indels 43; Gaps 10;

OY 5 GQPTQP---TYPVO-PPGNPVYPTLHLPOAPRYTDAPRAYSELYRSPVHPCATVPT 61
 DB 33 ggyppqgqppgyppgqppg---gyp-----gqap-----psay-----pgrtapgqaypgrt 84
 OY 62 MSAAFGASLYLPMASVAVGRLGSTITPMAYYPVGP-IVP-PGSTVLVEGCGYDAGARFGAG 119
 DB 85 pgaftpg-----qpgqgqaypypagqayp 120
 OY 120 ATAGNIP---PPPGCPNMAQL---AVMAGANLVYTORKN 155
 DB 121 tgppltpydpdpdgglmptr--mltlmglvkpnantivldfrrgn 162

RESULT 5
 ID W31855 standard; protein; 572 AA.
 XX
 AC W31855;
 XX
 DT 27-APR-1998 (first entry)
 XX
 DE Mycobacterium tuberculosis 55 kda protein.
 XX
 KM Tuberculosis; mycobacteria; infection; diagnosis;
 XX antimycobacterial; antibiotic; vaccine.
 OS
 XX Mycobacterium tuberculosis.
 PN W09741252-A2.
 XX
 PD 06-NOV-1997.
 XX
 PF 18-APR-1997; 97WO-EP01973.
 XX
 PR 29-APR-1996; 96DE-4017184.
 PR (GBFB) GBF GES BIOTECH FORSCHUNG GMBH.
 XX
 PI Espitla C, Honisch C, Moreno C, Singh M;
 DR WPI: 1997-549750/50.

DR N-PSDB: T93610.
 XX New DNA and related proteins or RNA derived from *M. tuberculosis*
 PT used for diagnosis of mycobacterial infections, monitoring
 PT vaccination and development of anti-mycobacterial agents
 PS
 XX Claim 11; Fig 16; 55pp; English.
 CC This novel 55 kDa protein is encoded by an open reading frame of
 CC a *Mycobacterium tuberculosis* DNA fragment (see T93610) containing
 CC polymorphic GC-rich sequences. Its amino acid sequence shows
 CC a high proline content, but there is no homology to any known
 CC proline-rich antigens of mycobacteria. Novel *M. tuberculosis*
 CC proteins (see W31851-57) are claimed. These can be produced as
 CC recombinant proteins, especially in bacterial, yeast, fungal or
 CC higher eukaryote host cells, and used for diagnosing tuberculosis
 CC and other mycobacterial infections in humans or animals. The
 CC claimed proteins can also be used for epidemiological studies, for
 CC monitoring vaccination, and for the development of vaccines and
 CC anti-mycobacterial drugs.
 CC
 SQ Sequence 572 AA;
 Query Match 12.1%; Score 111.5; DB 18; Length 572;
 Best Local Similarity 27.0%; Pred. No. 0.038;
 Matches 58; Conservative 8; Mismatches 90; Indels 59; Gaps 5;
 QY 2 NSKGQYPTQPTTYVQPGNVPYQTLHLQAPPYTDAPPVSELYRP----- 48
 DB 355 nspappappptppkllsappcpvpvppapnppappappappelpappdpptppvanspp 414
 QY 49 -----SFVHPGA-----ATVPTMSAARFGA-- 68
 DB 415 appappapppsalpfvnpappptppapksrpalpaappappvratltpppappapn 474
 QY 69 SLVLPMAQSVAVGPGSTIPMATYVPGPIYPPGCTVIVEGGYDAGARFGAGTAGNIPPP 128
 DB 475 smalpappdpdpilatltpappappplpmsppappl-----ppaappdpplltinqpps 529
 QY 129 PPGCPNMAQLAVMOGANVLVTORKGNFPMGSDG 163
 DB 530 pplapvpagaplapiingrvfark-nsligssg 563

RESULT 6
 W31852
 ID W31852 standard; Protein; 763 AA.
 AC W31852;
 XX
 DT 27-APR-1998 (first entry)
 DE *Mycobacterium tuberculosis* 74 kDa protein.
 XX
 KW Tuberculosis; mycobacteria; infection; diagnosis;
 KW antimycobacterial; antibiotic; vaccine.
 OS
 XX *Mycobacterium tuberculosis*.
 PN W09741252-A2.
 PD 06-NOV-1997.
 XX
 PF 18-APR-1997; 97WO-EP01973.
 XX
 PR 29-APR-1996; 96DE-4017184.
 XX
 PA (GBFB) GBF GES BIOTECH FORSCHUNG GMBH.
 XX
 PI Espitia C, Honisch C, Moreno C, Singh M;
 XX
 DR WPI; 1997-549750/50.

DR N-PSDB: T93610.
 XX
 XX New DNA and related proteins or RNA derived from *M. tuberculosis*
 PT used for diagnosis of mycobacterial infections, monitoring
 PT vaccination and development of anti-mycobacterial agents
 PS
 XX Claim 5; Fig 13; 55pp; English.
 CC This novel 74 kDa protein is encoded by an open reading frame of
 CC a *Mycobacterium tuberculosis* DNA fragment (see T93610) containing
 CC polymorphic GC-rich sequences. Its amino acid sequence shows
 CC a high proline content, but there is no homology to any known
 CC proline-rich antigens of mycobacteria. Novel *M. tuberculosis*
 CC proteins (see W31851-57) are claimed. These can be produced as
 CC recombinant proteins, especially in bacterial, yeast, fungal or
 CC higher eukaryote host cells, and used for diagnosing tuberculosis
 CC and other mycobacterial infections in humans or animals. The
 CC claimed proteins can also be used for epidemiological studies, for
 CC monitoring vaccination, and for the development of vaccines and
 CC anti-mycobacterial drugs.
 CC
 SQ Sequence 763 AA;
 Query Match 12.1%; Score 111.5; DB 18; Length 763;
 Best Local Similarity 27.0%; Pred. No. 0.052;
 Matches 58; Conservative 8; Mismatches 90; Indels 59; Gaps 5;
 QY 2 NSKGQYPTQPTTYVQPGNVPYQTLHLQAPPYTDAPPVSELYRP----- 48
 DB 546 nspappappptppkllsappcpvpvppapnppappappappelpappdpptppvanspp 605
 QY 49 -----SFVHPGA-----ATVPTMSAARFGA-- 68
 DB 606 appappappsalpfvnpappptppapksrpalpaappappvratltpppappapn 665
 QY 69 SLVLPMAQSVAVGPGSTIPMATYVPGPIYPPGCTVIVEGGYDAGARFGAGTAGNIPPP 128
 DB 666 smalpappdpdpilatltpappappplpmsppappl-----ppaappdpplltinqpps 720
 QY 129 PPGCPNMAQLAVMOGANVLVTORKGNFPMGSDG 163
 DB 721 pplapvpagaplapiingrvfark-nsligssg 754

RESULT 7
 R86913
 ID R86913 standard; Protein; 214 AA.
 AC R86913;
 XX
 DT 10-MAY-1996 (first entry)
 DE Cotton fiber-specific H6 protein.
 XX
 KW H6 gene; promoter; cotton; transgenic plant; fiber; immobilisation;
 KW enzyme; pesticide degradation; parathion hydrolase; extensin.
 OS
 XX *Gossypium hirsutum* strain Coker 312.
 FH
 FT Key Location/Qualifiers
 FT Peptide 1..25
 FT /label= Sig-peptide
 XX
 XX US5474925-A.
 PN
 XX
 PD 12-DEC-1995.
 XX
 PF 19-DEC-1991; 91US-0812233.
 XX
 PR 23-MAR-1994; 94US-0217327.
 XX
 PR 19-DEC-1991; 91US-0812233.
 XX
 DR

IGe; immunoglobulin; binding protein; allergy; allergic reaction;
absorbance; regulation; Immune response.

Rattus rattus.

US5260434-A.

09-NOV-1993.

23-APR-1985; 85US-0726250.

23-APR-1985; 85US-0726250.

21-SEP-1988; 88US-0247170.

10-JUL-1991; 91US-0728125.

(SCRI) SCRIPPS RES INST.

Liu F;

WPI: 1993-367951/46.

N-PSDB: Q50878.

DNA encoding Ige-binding protein - with repetitive sequence and

homology with Igg receptor

Claim 1; Figure 1; 20pp; English.

The Ige binding protein is used to regulate or absorb Ige. It is

useful particularly in the absorbance of Ige to ease allergic

reactions.

Sequence 262 AA:

Query Match 11.7%; Score 108; DB 14; Length 262;

Best Local Similarity 30.5%; Pred. No. 0.033;

Matches 51; Conservative 8; Mismatches 58; Indels 50; Gaps 12;

5 GQVP--TOP--TYPVQ--PPGNPVYPTLHPQAPYTDAPATSELYRSFVHPCAAIYPT 60

33 gypgsaygagaypgagpgp--yp-----gagpsaypgpigtgsay--Pgptagagyppt 84

61 MSAAFGASITLPMQSAVAVGLGSTRIMAYYPVGP--IYP--PGSTVLVEGGYDAGARFG 117

85 apgafpg-----gpgpgatpysapagypaapagypatpfg 120

118 AGATAGNIP--PPPGCPNMAQLAVMOGA-----NVLTQKGN 155

121 aptgpltpydmplpgvmpmr--mltllgtvkpnansitlnfkkn 165

RESULT 10

ID Y59724 standard; Protein: 725 AA.

XX Y59724;

XX 18-JAN-2000 (first entry)

XX Human normal ovarian tissue derived protein 1.

XX Human; ovary; screening; ovarian cancer; treatment.

XX Homo sapiens.

XX DEL9816395-A1.

XX 07-OCT-1999.

XX 03-APR-1998; 98DE-1016395.

XX 03-APR-1998; 98DE-1016395.

(META-) METAGEN GES GENOMFORSCHUNG MBH.

Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;

WPI: 1999-552352/47.

N-PSDB: Z41222.

Nucleic acid sequences potentially useful in diagnosis or therapy of

ovarian cancer

Claim 23; Page 215; 274pp; German.

This invention describes novel nucleic acid sequences that are highly

expressed in normal ovary tissue. Artificial chromosomes and cosmid

clones containing the sequences can be used as gene transfer vehicles.

The sequences can be used to produce DNA fragments containing full-length

genes. Host cells transformed with the sequences can be used to produce

polypeptides or polypeptide fragments, which can be used to screen phage

displays for polypeptides that bind to them, or as tools for identifying

agents active against ovarian cancer, or to prepare medicaments for

treating ovarian cancer. The cDNA sequences can be used to obtain genomic

genes, their promoters, enhancers, silencers, exon structures, intron

structures and their splice variants. Y59724-Y59837 represent protein

CC fragments encoded by the cDNA sequences represented in Z41222-Z41324

CC which are derived from normal human ovarian tissue.

Sequence 725 AA:

Query Match 11.7%; Score 108; DB 20; Length 725;

Best Local Similarity 28.1%; Pred. No. 0.098; 67; Indels 52; Gaps 1;

Matches 52; Conservative 14; Mismatches 67; Indels 52; Gaps 1;

2 NSKGQVPTOPTVOPPPGNPVYPTLHPQAPYTDAP-----AVSELYRSFVHPCAA 56

344 naagqlptspgh-----mhtqyppypgpgqpgqpgpigtgsanyrpg-----qp 390

57 TVPTMSAAPP-----GASLY-----LPMQSAVAVGLGSTRIMAYYPVGPPIYPGSTVL 105

391 vaptlnaypntpylssasysyqsllyaaqbhassptss--patsfp-----ppps----- 440

106 VEGGYDAGARF--GAGA-----TAGNIPPPPGCPNMAQLAVMOGANVLTQKGNFEM 158

441 -----sgastfngpgagpappssasayalppptgtlpaaselpsagqtengsldqgpmle 494

159 GGS DG 163

495 qpqng 499

RESULT 11

ID Y82707 standard; Protein: 1181 AA.

XX Y82707;

XX 11-AUG-2000 (first entry)

XX Human apoptosis related protein ABP125 SEQ ID NO.4.

XX Human; apoptosis related protein; apoptosis linked gene-2; ALC-2;

XX ABP125; ABP130; death trap; cancer; diagnosis; antibody.

XX Homo sapiens.

XX JP2000083672-A.

XX 28-MAR-2000.

XX 11-SEP-1998; 98JP-0258165.

XX 11-SEP-1998; 98JP-0258165.

PA (IGAK-) IGAKU SEIBUTSUGAKU KENKYUSHO KK.
 XX
 DR WPI: 2000-353336/31.
 DR N-PSDB: A29637.
 XX

XX An apoptosis-related protein
 PS
 XX
 PS Claim 1: Page 19-23; 39pp; Japanese.

XX The present sequence represents a human apoptosis related protein. Also
 CC described is a method for screening a compound having activity of
 CC inhibiting the combination between the an apoptosis related protein of
 CC the present invention and ALG-2 (apoptosis linked gene-2) protein
 CC including: (a) a step of contacting the apoptosis related protein with
 CC ALG-2 protein in the presence of a compound to be tested; (b) a step of
 CC detecting the combination between the above protein and ALG-2 protein;
 CC and (c) a step of selecting the compound having activity of inhibiting
 CC the combination between the apoptosis related protein and ALG-2 protein;
 CC a compound which has activity of inhibiting the combination between the
 CC apoptosis related protein and ALG-2 protein and can be isolated by the
 CC above method. An antibody from the present invention and the compound
 CC can be used for the elucidation of the molecular mechanism of apoptosis,
 CC and for the diagnosis and treatment of diseases such as cancers.
 CC
 XX

SQ Sequence 1181 AA:

Query Match 11.7%; Score 108; DB 21; Length 1181;
 Best Local Similarity 29.7%; Pred. No. 0.17; Mismatches 52; Indels 52; Gaps 10;
 Matches 49; Conservative 12;

OY 2 NSKQPTQPTTPYVOPGPNPVYPTLHLPOAPPYTDAPP-----AYSELYRPSFVHPGAA 56
 Db 813 naesqqltspgh-----mhtqvpypypqbpqypqfgygsamytpq-----qp 859
 OY 57 TVPTMSAAPP-----GASLY-----LPMQSVAVGPIGSTITPMAYYPVGPITPPGSTVL 105
 Db 860 vappptsnaypntlpylssasysytcgsgqlyaqhgasspts--patsf--ppps--- 909
 OY 106 VEGGYDAGARF---GAGA---TAGNIPPPPPGCPNMAOLAVMQ 143
 Db 910 -----sgasfghgpgpdpssasayalpqtgtclpaaseelpasq 948

RESULT 12

Y82708
 ID 182708 standard; Protein; 1220 AA.

AC Y82708;

DT 11-ANG-2000 (first entry)

DE Human apoptosis related protein ABP130 SEQ ID NO:6.

KW Human; apoptosis related protein; apoptosis linked gene-2; ALG-2;
 KW ABP32; ABP125; ABP130; death trap; cancer; diagnosis; antibody.
 XX

OS Homo sapiens.

PN JP2000083672-A.

PD 28-MAR-2000.

PF 11-SEP-1998; 98JP-0258165.

PR 11-SEP-1998; 98JP-0258165.

PA (IGAK-) IGAKU SEIBUTSUGAKU KENKYUSHO KK.

DR WPI: 2000-353336/31.

DR N-PSDB: A29638.
 PT An apoptosis-related protein

XX
 PS Claim 1: Page 30-35; 39pp; Japanese.
 XX

XX The present sequence represents a human apoptosis related protein. Also
 CC described is a method for screening a compound having activity of
 CC inhibiting the combination between the an apoptosis related protein of
 CC the present invention and ALG-2 (apoptosis linked gene-2) protein
 CC including: (a) a step of contacting the apoptosis related protein with
 CC ALG-2 protein in the presence of a compound to be tested; (b) a step of
 CC detecting the combination between the above protein and ALG-2 protein;
 CC and (c) a step of selecting the compound having activity of inhibiting
 CC the combination between the apoptosis related protein and ALG-2 protein;
 CC a compound which has activity of inhibiting the combination between the
 CC apoptosis related protein and ALG-2 protein and can be isolated by the
 CC above method. An antibody from the present invention and the compound
 CC can be used for the elucidation of the molecular mechanism of apoptosis,
 CC and for the diagnosis and treatment of diseases such as cancers.
 CC
 XX

SQ Sequence 1220 AA:

Query Match 11.7%; Score 108; DB 21; Length 1220;
 Best Local Similarity 29.7%; Pred. No. 0.17; Mismatches 52; Indels 52; Gaps 10;
 Matches 49; Conservative 12;

OY 2 NSKQPTQPTTPYVOPGPNPVYPTLHLPOAPPYTDAPP-----AYSELYRPSFVHPGAA 56
 Db 852 naesqqltspgh-----mhtqvpypypqbpqypqfgygsamytpq-----qp 898
 OY 57 TVPTMSAAPP-----GASLY-----LPMQSVAVGPIGSTITPMAYYPVGPITPPGSTVL 105
 Db 899 vappptsnaypntlpylssasysytcgsgqlyaqhgasspts--patsf--ppps--- 948
 OY 106 VEGGYDAGARF---GAGA---TAGNIPPPPPGCPNMAOLAVMQ 143
 Db 949 -----sgasfghgpgpdpssasayalpqtgtclpaaseelpasq 987

RESULT 13

R12531
 ID R12531 standard; Protein; 264 AA.

AC R12531;

DT 12-SEP-1991 (first entry)

DE Mac2.16 expression product.

KW HMEBP; leishmaniasis; Mouse Mac-2; laminin.

OS Mus musculus.

PN W09108290-A.

PD 13-JUN-1991.

PF 29-NOV-1990; 90MO-US06948.

PR 14-SEP-1990; 90US-0582628.

PR 30-NOV-1989; 89US-0444195.

PA (GEHO-) GEN HOSPITAL CORP.

DR WPI: 1991-193196/26.

DR N-PSDB: Q12207.

PT Recombinant gene encoding human macrophage carbohydrate -

PT Ige-binding protein and antibody used to treat, diagnose and
 PT prevent e.g. inflammatory bowel disorder, leishmaniasis, hayfever
 and bronchial

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 11:50:42 : Search time 34.57 Seconds
(without alignments)
87.266 Million cell updates/sec

Title: US-09-407-430-1

Perfect score: 921
Sequence: 1 MNSKQYPTQPTVPVPGN.....VTQRKNFPMGSDGCTIW 168

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/PCRTUS.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/Backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	115.5	12.5	264	1	US-08-562-311-4 Sequence 4, Appl
2	114.5	12.4	574	3	US-09-275-400-6 Sequence 6, Appl
3	110	11.9	214	1	US-08-217-327-4 Sequence 4, Appl
4	108	11.7	262	3	US-08-946-914-14 Sequence 14, Appl
5	105	11.4	259	3	US-08-476-509B-29 Sequence 29, Appl
6	103.5	11.2	902	1	US-08-396-479B-6 Sequence 6, Appl
7	103.5	11.2	902	1	US-08-818-823-6 Sequence 6, Appl
8	103.5	11.1	202	5	5202236-36 Patent No. 5202236
9	102.5	11.1	203	5	5202236-1 Patent No. 5202236
10	101.5	11.0	1248	2	US-09-080-897-2 Sequence 2, Appl
11	101.5	10.9	2441	1	US-08-194-468-2 Sequence 2, Appl
12	100.5	10.9	2441	3	US-08-961-739-2 Sequence 2, Appl
13	100	10.9	1336	2	US-08-231-193A-58 Sequence 58, Appl
14	100	10.9	1336	2	US-08-486-273A-58 Sequence 58, Appl
15	100	10.9	1336	3	US-08-940-086A-58 Sequence 58, Appl
16	99.5	10.8	330	3	US-08-642-255-32 Sequence 32, Appl
17	99.5	10.8	408	1	US-07-609-716-65 Sequence 65, Appl
18	99.5	10.8	408	1	US-08-475-411A-65 Sequence 65, Appl
19	98.5	10.7	652	5	5202236-13 Patent No. 5202236
20	98.5	10.7	744	5	5202236-25 Patent No. 5202236
21	98	10.6	546	1	US-08-498-168-10 Sequence 10, Appl
22	96	10.4	677	1	US-08-188-582-13 Sequence 13, Appl
23	96	10.4	677	1	US-08-646-715-13 Sequence 13, Appl
24	96	10.4	2414	4	US-08-227-536-2 Sequence 2, Appl
25	96	10.4	2414	4	PCT-US95-04682-2 Sequence 2, Appl
26	94.5	10.3	331	5	5202236-37 Patent No. 5202236
27	94.5	10.3	331	5	5202236-37 Patent No. 5202236
28	93.5	10.2	466	3	US-08-526-136-13 Sequence 13, Appl

29	93.5	10.2	503	3	US-08-526-136-2 Sequence 2, Appl
30	93.5	10.2	505	3	US-08-526-136-4 Sequence 4, Appl
31	92.5	10.0	267	3	US-08-301-162-16 Sequence 16, Appl
32	92	10.0	1719	2	US-08-459-568-4 Sequence 4, Appl
33	92	10.0	1719	2	US-08-399-411-4 Sequence 4, Appl
34	92	10.0	1719	3	US-08-516-859A-4 Sequence 4, Appl
35	91.5	9.9	129	5	5202236-5 Patent No. 5202236
36	91.5	9.9	267	1	US-07-706-872-3 Sequence 3, Appl
37	91.5	9.9	504	3	US-09-219-849-3 Sequence 3, Appl
38	91.5	9.9	561	1	US-08-642-255-52 Sequence 52, Appl
39	91.5	9.9	720	3	US-09-219-849-4 Sequence 53, Appl
40	91.5	9.9	777	1	US-08-642-255-53 Sequence 53, Appl
41	91.5	9.9	1315	3	US-08-899-595-3 Sequence 3, Appl
42	90.5	9.8	357	1	US-07-609-716-66 Sequence 66, Appl
43	90.5	9.8	357	1	US-08-642-255-33 Sequence 33, Appl
44	90.5	9.8	357	3	US-08-475-411A-66 Sequence 66, Appl
45	90.5	9.8	1255	2	US-09-080-897-4 Sequence 4, Appl

ALIGNMENTS

```
RESULT 1
US-08-562-311-4
; Sequence 4, Application US/08562311
; Patent No. 5801002
GENERAL INFORMATION:
APPLICANT: RAZ, AVRAHAM
TITLE OF INVENTION: A METHOD OF DETERMINING THE PROBABILITY
OF METASTASIS IN A CELL SAMPLE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dykema Gossett
STREET: STE 505 N. Woodward
CITY: Bloomfield Hills
STATE: MI
COUNTRY: U.S.
ZIP: 48304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/562,311
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/188,225
FILING DATE:
APPLICATION NUMBER: US 07/681,242
FILING DATE: 04-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/294,249
FILING DATE: 01-JUN-1989
ATTORNEY/AGENT INFORMATION:
NAME: KELLY, ROBERT L.
REGISTRATION NUMBER: 31,843
REFERENCE/DOCKET NUMBER: 61,686-
TELECOMMUNICATION INFORMATION:
TELEPHONE: 810-540-0849
TELEFAX: 810-540-0763
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 264 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-562-311-4
Query Match 12.5%, Score 115.5; DB 1: Length 264;
Best Local Similarity 32.5%, Pred. No. 0.0027;
```


TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 262 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-946-914-14

Query Match
Best Local Similarity 11.7%; Score 108; DB 3; Length 262;
Matches 51; Conservative 8; Mismatches 58; Indels 50; Gaps 12;

Db 33 GGYPGASVPGAYPGQAPGCG--Y-----GQAPPSAVPGPTGSAV-PCPTAGAVPGPT 84

QY 5 GQYPT--TOP-TYVQ--PCGNPVYPTLHLQAPPYTDAPPVSELYRPSFVHPGAAVTP 60

Db 61 MSAAFGASLYLPMASVAVGRLGSTITPAAVYVGP--YTP--PGSTVLYEGGIDAGARG 117

QY 85 AFGAFG-----PGGPGAVPSAPGAVPSAPGAVPATGPG 120

QY 118 AGATAGNIP---PPPCPPNMAQLAVMOGA-----NYLVYQKCN 155

Db 121 APTGPLTVPYDMPLPGCVMPR--MLTITIGYKPMANSTILNKKCN 165

RESULT 5
US-08-476-509B-29
Sequence 29, Application US/08476509B
Patent No. 6034212

GENERAL INFORMATION:
APPLICANT: SUDOL, MARIUS
APPLICANT: PEER, BORK
TITLE OF INVENTION: A SH2 DOMAIN ASSOCIATED PROTEIN, A
TITLE OF INVENTION: SIGNALING DOMAIN THEREIN, NUCLEIC ACIDS ENCODING THE
TITLE OF INVENTION: PROTEIN AND THE DOMAIN, AND DIAGNOSTIC AND THERAPEUTIC USES
TITLE OF INVENTION: THEREOF
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESS: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/476-509B
FILING DATE: 01-DEC-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-101 CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521

INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 259 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: peptide
HYPOTHETICAL: NO
US-08-476-509B-29

Query Match
Best Local Similarity 11.4%; Score 105; DB 3; Length 259;
Matches 38; Conservative 8; Mismatches 41; Indels 54; Gaps 8;

QY 3 SKGQYPTQPTTYVQPP-----GNPVYPTLHLQAPPYTDAPPVSELYRPSFVHPGAA 56

Db 158 ANGMYPCCPPGYPPPPPEFYAGPPMDGAMGVQPP-----YPCPM 202

QY 57 TYPTMSAARPGASLYLPMASVAVGRLGSTITPAAVY-PVGP--YTPGSTVLYEGGIDAG 113

Db 203 EPPVSGPSAPAT---PAAEAKAAEAAAS---AYVPCNPINHVMPYTSQ----- 244

QY 114 ARFGAGATAGNIPPPPPGCP 134

Db 245 -----PPPPPYTP 253

RESULT 6
US-08-396-479B-6
Sequence 6, Application US/08396479B
Patent No. 5612455

GENERAL INFORMATION:
APPLICANT: HOEY, Timothy
TITLE OF INVENTION: NUCLEAR FACTORS AND BINDING ASSAY
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESS: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/396,479B
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A.
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-59450-1/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 494-8700
TELEFAX: (415) 494-8771
TELEX: 210 277299

INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 902 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-396-479B-6

Query Match
Best Local Similarity 11.2%; Score 103.5; DB 1; Length 902;
Matches 47; Conservative 12; Mismatches 68; Indels 61; Gaps 8;

QY 8 PTOPTYPQPGNPVYPTLHLQAPPYTDAPPVSELYRPSFVHPGAAVTP---TMSA 63

Db 717 PPRPPYPPPHEDPAC-ETPILSEFGY-GMPPLYPOTGPPPSYRPGURMPETRGTTGC 774

QY 64 AFGASLYLPMASVAVGRLGSTITP----- 89

Db 775 AOPPAVSFLPRPPSPDYGGGSSFLGLPFPSPAPRRPPLASPLLEGPFSQSDVHP 834
QY 90 ----AYYPVGPPIYPPGSTVLVE----GGYDAGARFGA---GATAGNIP-----PP 128
Db 835 LPABGYKVGKVGCGGAGAEBOEKSRGSGSFRDVSPIQGITLLEVESEIIGRDLGCFPA 894
QY 129 PPGCPNPA 136
Db 895 PPGCEPPA 902

RESULT 7

US-08-818-823-6
; Sequence 6, Application US/08818823
; Patent No. 5708158
; GENERAL INFORMATION:
; APPLICANT: HOEY, Timothy
; TITLE OF INVENTION: NUCLEAR FACTORS AND BINDING ASSAY
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; City: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,823
; FILING DATE: 14-MAR-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/396,479
; FILING DATE: 02-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-59450-1/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 210 277299
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 902 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-818-823-6

Query Match 11.2%; Score 103.5; DB 1; Length 902;
Best Local Similarity 25.0%; Pred. No. 0.13; Mismatches 68; Indels 61; Gaps 8;
Matches 47; Conservative

QY 8 PTQPTYPVOPGPNVYPTQLHL-PQAPPYTDAPPAYSELYRPSFVHPGATVTP-----TMSA 63
Db 717 PPRPPYPSYPRHEDPAC-ETPYLSEGFY-GMPPLYPQTGPPSYKRLGMPFETRTGTGC 774
QY 64 AFGASLTPMA-QSVANGPLGSTIPM----- 89
Db 775 AOPPAVSFLPRPPSPDYGGGSSFLGLPFPSPAPRRPPLASPLLEGPFSQSDVHP 834
QY 90 ----AYYPVGPPIYPPGSTVLVE----GGYDAGARFGA---GATAGNIP-----PP 128
Db 835 LPABGYKVGKVGCGGAGAEBOEKSRGSGSFRDVSPIQGITLLEVESEIIGRDLGCFPA 894
QY 129 PPGCPNPA 136
Db 895 PPGCEPPA 902

Db 895 PPGCEPPA 902

RESULT 8
5202236-36
; Patent No. 5202236
; APPLICANT: MAUGH, KATHY J.; ANDERSON, DAVID M.; STRAUSBERG,
; SUSAN L.; MCCANDLISS, RUSS; WEI, TENA; FILIPULA, DAVID
; TITLE OF INVENTION: METHOD OF PRODUCING BROADHESIVE
; PROTEIN
; NUMBER OF SEQUENCES: 39
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/528,762
; FILING DATE: 25-MAY-1990
; APPLICATION NUMBER: 82,456
; FILING DATE: 07-AUG-1987
; APPLICATION NUMBER: 933,945
; FILING DATE: 24-NOV-1986
; APPLICATION NUMBER: 650,128
; FILING DATE: 13-SEP-1984
; SEQ ID NO: 36
; LENGTH: 202
5202236-36

Query Match 11.1%; Score 102.5; DB 5; Length 202;
Best Local Similarity 26.4%; Pred. No. 0.029; Mismatches 59; Indels 35; Gaps 6;
Matches 37; Conservative

QY 7 YPTQPTYPVOPGPNVYPTQLHL-PQAPPYTDAPPAYSELYRPSFVHPGATVTPMSA 65
Db 30 YKAPSYATYKAKPSYPTTYKAKPSYPTTYKAKPSYPTTYKAKPSYPTTYKAK 89
QY 66 PG-----ASLYLPMASVAVGFLGSTIMAYVPVGPPIYPPGSTVLVEGGYDAGARFGA 118
Db 90 PSYPTTYKPKPSYPTTYKSKSTIP-SSYKPKKTIY--PTYKPKLT----- 131
QY 119 GATAGNIP---PPGCPN 135
Db 132 -----YPTTYKPKPSYPS 145

RESULT 9
5202236-1
; Patent No. 5202236
; APPLICANT: MAUGH, KATHY J.; ANDERSON, DAVID M.; STRAUSBERG,
; SUSAN L.; MCCANDLISS, RUSS; WEI, TENA; FILIPULA, DAVID
; TITLE OF INVENTION: METHOD OF PRODUCING BROADHESIVE
; PROTEIN
; NUMBER OF SEQUENCES: 39
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/528,762
; FILING DATE: 25-MAY-1990
; APPLICATION NUMBER: 82,456
; FILING DATE: 07-AUG-1987
; APPLICATION NUMBER: 933,945
; FILING DATE: 24-NOV-1986
; APPLICATION NUMBER: 650,128
; FILING DATE: 13-SEP-1984
; SEQ ID NO: 1
; LENGTH: 203
5202236-1

Query Match 11.1%; Score 102.5; DB 5; Length 203;
Best Local Similarity 26.4%; Pred. No. 0.029; Mismatches 59; Indels 35; Gaps 6;
Matches 37; Conservative

QY 7 YPTQPTYPVOPGPNVYPTQLHL-PQAPPYTDAPPAYSELYRPSFVHPGATVTPMSA 65
Db 31 YKAPSYATYKAKPSYPTTYKAKPSYPTTYKAKPSYPTTYKAKPSYPTTYKAK 90
QY 66 PG-----ASLYLPMASVAVGFLGSTIMAYVPVGPPIYPPGSTVLVEGGYDAGARFGA 118

ORGANISM: Mus
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1) ... (2441)
OTHER INFORMATION: xaa - Any Amino Acid
US-08-961-739-2

Query Match 10.9%; Score 100.5; DB 3; Length 2441;
Best Local Similarity 28.3%; Pred. No. 0.77;
Matches 43; Conservative 9; Mismatches 51; Indels 49; Gaps 7;

QY 3 SKGQIPPTPTVPQPGNPVYPTQLHLPQAPPTDAPPAVSELYRPSFVHGATVPTMS 62
DB 822 SQGQ---EPGALPPLNMLAPQASQLP---CPPTQSP-----LHP---TPPPAS 863
QY 63 AAFGASLYLPMAOSVANGPLGSTIPMAVYVGPITPPGSTVLVEGVDAGARFGAGATA 122
DB 864 TAAGMPSLQHTAP-----GMTPPQPAATPTQSTP-----VSS 896
QY 123 GNIPPPPGCPNNAQL---AVMGANVLVT 150
DB 897 GQTPPTPGSVSAAGTOSTPTVQAAQAQVT 928

RESULT 13
US-08-231-193A-58
Sequence 58, Application US/08231193A
Patent No. 5849895

GENERAL INFORMATION:
APPLICANT: Daggett, Lorie P.
APPLICANT: Ellis, Steven B.
APPLICANT: Liaw, Chen W.
APPLICANT: Lu, Chin-Chun
TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR
TITLE OF INVENTION: SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR
NUMBER OF SEQUENCES: 63
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Brown, Martin, Haller & McClain
STREET: 1660 Union Street
CITY: San Diego
STATE: CA
COUNTRY: U.S.A.
ZIP: 92101-2926

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/231,193A
FILING DATE: 20-APR-1994
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/052,459
FILING DATE: 20-APR-1993
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 6362-9383
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-238-0999
TELEFAX: 619-238-0062

INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 1336 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-231-193A-58

Query Match 10.9%; Score 100; DB 2; Length 1336;
Best Local Similarity 27.1%; Pred. No. 0.43;
Matches 45; Conservative 9; Mismatches 74; Indels 38; Gaps 8;

QY 19 GNPVYPTQLHLPQAPPYT-----DAPPAYSELYRPSFVHGATVPTMSAEPGASLYL 72
DB 989 GRPLSPPAAGPQKPPASVFAIVRDKPEAP-----PAGAFPGFPSPAP----- 1033
QY 73 PMAOSVANGPLGSTIPMAVYVGPITP-----PGSTVLEGGVDAGARFGAGATAGNI 25
DB 1034 PAAATAVAGP--PLCRLAFEDESPAPARMPRSDPESQPL--GPAGAGAGTGAGGGA 1089
QY 126 PPPPGC---PPNAAQLAVMGANVLVTRKGNFPMGSDGTYTW 168
DB 1090 PAAPPPCFAPPCFYLDVQSPS---DSHDSLSLAGASLAGLDPW 1132

RESULT 14
US-08-486-273A-58
Sequence 58, Application US/08486273A
Patent No. 5985586

GENERAL INFORMATION:
APPLICANT: Daggett, Lorie P.
APPLICANT: Ellis, Steven B.
APPLICANT: Liaw, Chen W.
APPLICANT: Lu, Chin-Chun
TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR SUBUNITS, DNA
TITLE OF INVENTION: ENCODING SAME AND USES THEREFOR
NUMBER OF SEQUENCES: 63
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Brown, Martin, Haller & McClain
STREET: 1660 Union Street
CITY: San Diego
STATE: CA
COUNTRY: U.S.A.
ZIP: 92101-2926

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/486,273A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/231,193
FILING DATE: 20-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 6362-9383B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-238-0999
TELEFAX: 619-238-0062

INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 1336 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-486-273A-58

Query Match 10.9%; Score 100; DB 2; Length 1336;
Best Local Similarity 27.1%; Pred. No. 0.43;
Matches 45; Conservative 9; Mismatches 74; Indels 38; Gaps 8;

QY 19 GNPVYPTQLHLPQAPPYT-----DAPPAYSELYRPSFVHGATVPTMSAEPGASLYL 72
DB 989 GRPLSPPAAGPQKPPASVFAIVRDKPEAP-----PAGAFPGFPSPAP----- 1033

QY 73 PMAQSVAVGELGSTITPMAYYPVGPITYP-----PGSTVLVEGGYDAGARFGAGATAGNI 125
DB 1034 PAAATAVAVGP--PLCRLAFEDSEPPAPARMPRSDPESQPL--GPGAGGAGGTGGAGGCA 1089
QY 126 PPPPPGC---PPNAAQLAVMOGANVLTORKGNFEMGSGSGGYTTI 168
DB 1090 PAAPPPCFAPPPCFYLDVDSPPS---DSRDSLSLAGASLAGIDPW 1132

RESULT 15

US-08-940-086A-58
Sequence 58, Application US/08940086A
Patent No. 6111091
GENERAL INFORMATION:
APPLICANT: Daggett, Lorrie P.
APPLICANT: Ellis, Steven B.
APPLICANT: Liaw, Chen W.
APPLICANT: Lu, Chin-Chun
TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR
TITLE OF INVENTION: SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR
NUMBER OF SEQUENCES: 63
CORRESPONDENCE ADDRESS:
ADDRESSEE: Heller Ehtman White & McAuliffe
STREET: 4250 Executive Square, 7th Floor
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/940,086A
FILING DATE: 29-SEPT-97
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/231,193
FILING DATE: 20-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/052,449
FILING DATE: 20-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 24735-9383C
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 450-8400
TELEFAX: (619) 450-8499
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 1336 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-940-086A-58

Query Match 10.9%; Score 100; DB 3; Length 1336;
Best Local Similarity 27.1%; Pred. No. 0.43;
Matches 45; Conservative 9; Mismatches 74; Indels 38; Gaps 8;

QY 19 GNPVYPTQLHLPDAPRYT-----DAPPAYSFLYRPSFVHPGAATVPTMSAAPPASLYL 72
DB 989 GRPLSPPAQPPQKPPASTFAIVRDEPEAP---PAGAFPGFSPSPAD-----1033
QY 73 PMAQSVAVGELGSTITPMAYYPVGPITYP-----PGSTVLVEGGYDAGARFGAGATAGNI 125
DB 1034 PAAATAVAVGP--PLCRLAFEDSEPPAPARMPRSDPESQPL--GPGAGGAGGTGGAGGCA 1089
QY 126 PPPPPGC---PPNAAQLAVMOGANVLTORKGNFEMGSGSGGYTTI 168

DB 1090 PAAPPPCFAPPPCFYLDVDSPPS---DSRDSLSLAGASLAGIDPW 1132
Search completed: March 6, 2001, 11:52:11
Job time: 89 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 11:50:43 ; Search time 45.68 Seconds

(Without alignments)
249.722 Million cell updates/sec

Title: US-09-407-430-1

Sequence: 1 MNSKGOVPTQPTYPVQPGN.....VTORKGNFMGSDGGYTIW 168

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	133.5	14.5	620	2	hydroxyproline-rich
2	129	14.0	502	2	hypothetical prote
3	126	13.7	856	2	ELB-55kDa-associat
4	124	13.5	231	2	proline-rich prote
5	124	13.5	491	2	DNA-directed RNA p
6	124	13.5	1834	1	DNA-directed RNA p
7	123.5	13.4	2715	2	eyelid - fruit fly
8	123	13.4	625	2	DNA-directed RNA p
9	123	13.4	977	2	DNA-directed RNA p
10	121.5	13.2	977	2	DNA-directed RNA p
11	120	13.0	1126	2	DNA-directed RNA p
12	119	12.9	1541	2	T02831
13	118	12.8	1690	2	T31670
14	117.5	12.8	530	2	T48627
15	117.5	12.8	1896	1	RNPE2L
16	117	12.7	1859	1	A34092
17	117	12.7	1862	1	T29559
18	116.5	12.6	708	1	QOBE88
19	116.5	12.6	990	2	T14756
20	116.5	12.6	1553	2	T31436
21	116.5	12.6	1733	1	RNBY2L
22	115.5	12.5	264	2	A45983
23	114.5	12.4	574	1	A48501
24	113	12.3	306	2	T09067
25	113	12.3	1585	2	T31611
26	113	12.3	1605	2	T31435
27	112	12.1	346	2	S19129
28	111.5	12.1	237	2	S46964
29	111.5	12.1	467	2	A27677

30	111.5	12.1	1932	1	A28490
31	111.5	12.1	1970	1	S21054
32	111.5	12.1	1370	2	T38186
33	111	12.1	379	2	T05441
34	111	12.1	497	2	S22708
35	111	12.1	554	2	T49833
36	110.5	12.0	698	2	T32594
37	110	11.9	214	2	T09854
38	109.5	11.9	275	2	T51437
39	109.5	11.9	289	2	T26812
40	109.5	11.9	815	2	B56708
41	108.5	11.8	142	2	B41132
42	108.5	11.8	142	2	S21930
43	108.5	11.8	259	2	S01704
44	108.5	11.8	419	2	T29266
45	108	11.7	214	2	T10737

ALIGNMENTS

RESULT 1

S06733 hydroxyproline-rich glycoprotein precursor - common tobacco

C:Species: Nicotiana tabacum (common tobacco)

C:Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 21-Jul-2000

C:Accession: S06733

R:Keller, B.; Lamb, C.J.

Genes Dev. 3, 1639-1646, 1989

A:Title: Specific expression of a novel cell wall hydroxyproline-rich glycoprotein ge

A:Reference number: S06733; MUID:90128263

A:Accession: S06733

A:Molecule type: DNA

A:Residues: 1-620 <KE>

A:Cross-references: EMBL:X13885; NID:g19866; PIDN:CAA32090.1; PID:g19867

C:Superfamily: hydroxyproline-rich glycoprotein

C:Keywords: glycoprotein

Query Match 14.5%; Score 133.5; DB 2; Length 620;

Best Local Similarity 26.2%; Pred. No. 0.0064;

Matches 48; Conservative 10; Mismatches 50; Indels 75; Gaps 7;

OY	3	SKGOVPTQPTYPVQPGNVPQT-----LHL-----	PGAPPTTDAPPYSELTPPS 49
DB	220	SKHDPQPTNHNAPRTNHPSPDLRLPPSPRPQPPYSPPPAYAOSSPPS 279	
OY	50	FVN-----	-----LYLP 73
DB	280	PTYSPPPTTYSPPSPPTISPPRPAYSPSPPTPTTSPPPPPAYSPPTTYSPPPTTLP 339	
OY	74	MAOSAVAGPLGSTTPMAYVVGSP--IYPPGSTVLVEGGYDAGARCGATAGNIPPPPG 131	
DB	340	LPSS-----PTYSPPPPVYSPPPPSYSPPPPTLP-----	-----PPPPS 375
OY	132	CP 134	
DB	376	SPP 378	

RESULT 2

T08776 hypothetical protein DKFZP586D0920.1 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 13-Aug-1999

C:Accession: T08776

R:Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, May 1999

A:Reference number: Z16469

A:Accession: T08776

A:Molecule type: mRNA

A:Residues: 1-502 <MAN>

A:Cross-references: EMBL:AL050146


```
A:Reference number: 215380  
A:Accession: T04690  
  
Molecule type: DNA  
A.Residues: 1-1834 <BEV>  
A.Cross-references: EMBL:AL031986  
A.Experimental source: cultivar Columbia; BAC clone F4B14  
R.Nawrathl, C.; Schell, J.; Koncz, C.  
Mol. Gen. Genet. 223, 65-75, 1990  
A.Title: Homologous domains of the largest subunit of eucaryotic RNA polymerase II are c  
A.Reference number: S12071; MUID:g1080867  
A.Accession: S12071  
A.Molecule type: DNA  
A.Residues: 1-421,'S','423-'733','D','734'-1055,'R','1057'-1714,'SP'PSPV','1715-1834 <NAN1>  
A.Cross-references: EMBL:X52954; NID:g16504; PIDN:CNA37130.1; PDI:g16505  
A.Experimental source: cv. Columbia  
A.Note: The authors translated the codon AGC for residue 1755 as Arg  
A.Accession: S27346  
A.Molecule type: mRNA  
A.Residues: 510-'73','O','734'-1055,'R','1057'-1714,'SP'PSSPY','1715-1834 <NAM2>  
A.Experimental source: cv. Columbia  
R.Dietrich, M.A.; Prenger, J.P.; Gullfoyle, T.J.  
Plant Mol. Biol. 15, 207-223, 1990  
A>Title: Analysis of the genes encoding the largest subunit of RNA polymerase II in Arab  
A.Reference number: S11960; MUID:g1353869  
A.Accession: S11960  
A.Molecule type: DNA  
A.Residues: 1-116,'125'-192,'NSKEE','198'-297,'R','299'-302,'R','304'-400,'KE','403','YDCGHPRRPKTK'  
A.Cross-references: EMBL:X52954; NID:g16493; PIDN:CAA36735.1; PDI:g16494  
A>Note: The authors translated the codon CCT for residue 1083 as Ala  
C.Genetics:  
A.Gene: rpl1215, RPB1  
A.Map position: 4  
A.Intons: 28/'3; 123/'3; 218/'3; 325/'3; 411/'1; 440/'3; 552/'3; 648/'2; 734/'3; 1755/'2;  
A.Note: F4B14.'70  
C.Superfamily: human DNA-directed RNA polymerase II largest chain  
K.Words: DNA binding; nucleic acid transferase; phosphoprotein; tandem repeat; transcr  
F:65-109/Region: zinc finger CCCC motif  
F:1531-1812/Region: 7-residue repeats
```

Query Match 13.5%; Score 124; DB 1; Length 1834;

Best Local Similarity 27.1%; Pred. No. 0.1;

Matches 42; Conservative 17; Mismatches 72; Indels 24; Gaps 6;

Oy 8 PROOPT-EVOORGNVPYOQLHLPOAPPYTDAAPAYSELYNRSGAAATVTPTMSAAF 66
Db 1667 PSPPSYSPSPSPSPSPSPSP--YSPTSYPASVSPTSPQSPST-SPSYSPSPSPSPSPSNP 1723
Oy 67 GASLVLPMAOSVAAGVCPLGSTIPMA-----YYPCRTTPRSS-TVLVEGG 109
Db 1724 QSAKSFP---SLAYSFSNNKRISPAASPVSPTSPBNISPSPSPSPSPSPSPSPSPSPSPSP 1780
Oy 110 YDGARFGAGATTAGNIPIPPPPECPRPNMAQLAVMOG 144
Db 1781 YSSGASPDYSPSAGSYSPLLPGVSPSTGTGYMPHEG 1815

RESULT 7
T13049
eyelid - fruit fly (*Drosophila melanogaster*)
C.Species: Drosophila melanogaster
C.Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 13-Aug-1999
C.Accession: t13049
R.Treisman, J.E., Luk, A., Rubin, G.M., Heberlein, U.
submitted to the EMBL Data Library, March 1998
A.Reference number: Z17592
A.Accession: t13049
A.Status: preliminary; translated from GB/EMBL/DDBJ
A.Molecule type: mRNA
A.Residues: 1-2715 <TRE>
A.Cross-references: EMBL:AF053091; NID:g2981220; PID:g2981221; PIDN:AAC06254.1
C.Genetics:
A.Gene: eld

C:Function: could act as a transcription factor antagonistic to the wg pathway
C:Keywords: DNA binding

Query Match 13.4% ; Score 123.5; DB 2; Length 2715;
Best Local Similarity 31.0%; Pred. No. 0.17; Mismatches 81; Indels 33; Gaps 8;

MATCHES 53: Conservative 4; Mismatches 67; Indels 33; Gaps 8.

OY 1 MNKSGQPTPTPPVQ-----PGNPVPYPQTLLHPDAP--YTADAPPAYSELRYRSPFWIP 53
DB 63G MGCPSSPCACGPFPQQPQPQGPN-YPFR---POYPRGAATGAP----- 677
OY 54 GAATPVPMASARPGAISLVLPMAOSVAVGELGSTIMATVPYPIPPPSSTVLVEGYDAG 113
DB 67B ---PPPTSQAAGGGGNMSPCAQAGCYPERGRPHNTGTOPYQWVPSPDOTVPCGABCG 734
OY 11A AREGAGATAGNIPPDP-PCCPPNAQLAVMOGANLYL-TORKCNPFMGCS 161
DB 73S AMYGHNHVOCKTGFPPPPVVGVGPPPPGSGSRPLNYLKQHJHKKGYY-CGS 783

RESULT 8

T07796
DNA-directed RNA polymerase (EC 2.7.7.6).largest chain - soybean (fragment)
C:Species: Glycine max (soybean)
C>Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 08-Oct-1999
C:Accession: T07796
R:Dietrich, M.A.; Prenger, J.P.; Guilfoyle, T.J.
Plant Mol. Biol. 15, 207-223, 1990
A>Title: Analysis of the genes encoding the largest subunit of RNA polymerase II in A
A:Reference number: S11960; MUID:91355869
A:Accession: T07796
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-625 <DIE>
A:Cross-references: EMBL:X52492; NID:g18731; PIDN:CAA36733.1; PUD:g18732
A:Experimental source: BL strain var.Wayne
C:Genetics:
A:Gene: RPBI-B1
A>Note: Intron positions not resolved (Incomplete sequence)
C:Superfamily: human DNA-directed RNA polymerase II largest chain
C:Keywords: DNA binding; nucleotidylyltransferase; transcription

Query Match 13.4% ; Score 123; DB 2; Length 625;
Best Local Similarity 27.7%; Pred. No. 0.039;
MATCHES 41: Conservative 12; Mismatches 67; Indels 28; Gaps 5;

OY 3 SKGOIPTOPTPYVQPC---NPVYPOT-----LHLPAAPPYTDAPPAYSSELYRPS 49
DB 39S SPGVSPTSPYSRSSPSCGTSPPAYSPSYSPYSPTSPYSPTSPYSPTSP-SFS 453
OY 50 FVHGAAIVPIMSAAFPCASLIYLPAQSVA-VGPLGSTIIMPAYIPVGIIRPGSTVLVEG 108
DB 45A YSPTPSPAISPSPASPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP---- 508
OY 109 GYDAGARPAGATAGNIPPPCCPPNA 136
DB 509 -----SYSPTSPAVSPSPSPSPSPS 528

RESULT 9

S14181
DNA-directed RNA polymerase (EC 2.7.7.6) .largest chain (isoform B1) - soybean (fragme
C:Species: Glycine max (soybean)
C>Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 12-Sep-1997
C:Accession: S14181
R:Dietrich, M.A.; Prenger, J.P.; Guilfoyle, T.J.
Plant Mol. Biol. 15, 207-223, 1990
A>Title: Analysis of the genes encoding the largest sbsubit of RNA polymerase II in A
A:Reference number: S11960; MUID:91355869
A:Accession: S14181

Tue Mar 6 12:02:32 2001

us-09-407-430-1.rpt

Page 6

QY 97 IYPPGSTVIVE 107
| | | |
Db 1876 AYSPSSPTFEE 1886

Search completed: March 6, 2001, 11:53:02
Job time: 139 sec

32	104.3	11.3	503	1	ANXA_RABIT
33	104	11.3	268	1	CDY1_MOUSE

45	102	11.1	464	1	SP62_HUMAN
44	102	11.1	379	1	YPR5_CAEEL
43	102.5	11.1	1281	1	PER1_MOUSE
42	102.5	11.1	1132	1	BAT3_HUMAN
41	102.5	11.1	304	1	GAT1_CHICK
40	103	11.2	639	1	PAB2_ARATH
39	103.5	11.2	1183	1	DRPL_RAT
38	103.5	11.2	902	1	NFC4_HUMAN
37	103.5	11.2	872	1	FPI_MITCO
36	103.5	11.2	751	1	FPI_MYGA
35	104	11.3	485	1	SP62_MOUSE
34	104	11.3	302	1	PIX3_MOUSE
33	104	11.3	35	1	035160
32	104	11.3	35	1	062200
31	104	11.3	35	1	027409
30	104	11.3	35	1	027409
29	104	11.3	35	1	027409
28	104	11.3	35	1	027409
27	104	11.3	35	1	027409
26	104	11.3	35	1	027409
25	104	11.3	35	1	027409
24	104	11.3	35	1	027409
23	104	11.3	35	1	027409
22	104	11.3	35	1	027409
21	104	11.3	35	1	027409
20	104	11.3	35	1	027409
19	104	11.3	35	1	027409
18	104	11.3	35	1	027409
17	104	11.3	35	1	027409
16	104	11.3	35	1	027409
15	104	11.3	35	1	027409
14	104	11.3	35	1	027409
13	104	11.3	35	1	027409
12	104	11.3	35	1	027409
11	104	11.3	35	1	027409
10	104	11.3	35	1	027409
9	104	11.3	35	1	027409
8	104	11.3	35	1	027409
7	104	11.3	35	1	027409
6	104	11.3	35	1	027409
5	104	11.3	35	1	027409
4	104	11.3	35	1	027409
3	104	11.3	35	1	027409
2	104	11.3	35	1	027409
1	104	11.3	35	1	027409

RESULT	1	
EXTN_TOBAC		
ID	EXTN_TOBAC	STANDARD;
		PRT;
		620 AA

DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-MAR-1992 (Rel. 21, Last annotation update)
 DE EXTENSIN PRECURSOR (CELL WALL HYDROXYPROLINE-RICH GLYCOPROTEIN).
 GN HRPNT3.
 OS Nicotiana tabacum (Common tobacco).
 CC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta
 CC Magnoliophyta; eudicotyledons; core eudicots; asterids; euasterids
 CC Solanales; Solanaceae; Nicotiana.
 [1]

RA "Specific expression of a novel cell wall hydroxyproline-rich
RT glycoprotein gene in lateral root initiation.";
RT Genes Dev. 3:1639-1646(1989).
RX
RX MEDLINE=90128263; PubMed=2612909;
RC
RC SEQUENCE FROM N.A.
RP STRAIN=CV. XANTHI. TISSUE=LEAF;
RP MEDLINE=90128263; PubMed=2612909;

```

CC -|- FUNCTION: HAS A SPECIALIZED STRUCTURAL FUNCTION, POSSIBLY IN
CC THE MECHANICAL PENETRATION OF THE CORTEX AND EPIDERMIS OF THE
CC MAIN ROOT.

```

CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.
CC -1- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE
CC SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN
CC GLYCOSYLATED.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcements>
CC or send an email to license@sib-ch.ch).

CC
DR EMBL: X13885, CAA32090.1, -.
PIR: S06733; S06733.
KW Repeat; Cell wall; Glycoprotein; Signal; Structural protein;
uncharacterized

SEQUENCE	620 AA.	65406 MW,	641DD2278ABD8524 CRG64.
DOMAIN	498	600	CONTAINS THE SER-PRO(4) REPEATS.
DOMAIN	205	630	3 X APPROXIMATE TANDEM REPEATS.
REPEAT	236	242	2.
REPEAT	229	235	1.
REPEAT	242	229	2 X 7 AA TANDEM REPEATS OF T-H-R-H-A-P-P
DOMAIN	148	151	H-A-P-P.
REPEAT	70	73	H-A-P-P.
CHAIN	1	630	EXTENSIN.
SIGNAL	1	?	
HYDROXYLATION.			

Query Match	14.58;	Score 133.5;	DB 1;	Length 620;
Best Local Similarity	26.28;	Pred. No. 0.013;		
Matches 48;	Conservative 10;	Mismatches 50;	Indels 75;	Gaps 7;

```
QY 3 SKGOYPTQPTVYVOPPCGNPVYPTQ-----LHL-----PQAPPYTDAPPAYSELYRPS 49
D 220 SRGHQPPPTPHRAAPPTPHRAAPTHQPSPLRLHPPSPRRQPPQPTYSPPPPAYAQSPPS 279
QY 50 FVH-----PCGATVPTMSAFAFGAS-----LYLP 73
D 280 PTYSPPPPPTYSPPPPPTYSPPPPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 339
QY 74 MAOSVAVGPGTSTIPMAYYPVGP--IYPPGSTVLVEGGYDAGARFAGAGATAGNITPPPPPG 131
D 340 LPSS-----PTYSPPPPYSPPPPPSPPTPTYL-----PPPPPS 375
QY 132 CPP 134
D 376 SPP 378

RESULT 2
CBPA_DICTDI STANDARD; PRT; 467 AA.
ID CBPA_DICTDI
AC P35085;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CALCIUM-BINDING PROTEIN.
GN CBPA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Dictyostelida; Dictyostelium.
RC SEQUENCE FROM N.A.
RP STRAIN-AX2;
RA Weinington R., Greenwood M., Tsang A.;
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
CC -1- SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U03413; AAA03471.1; -.
DR DICTYDB; DD01055; CBPA.
DR INTERPRO: IPR002048; -.
DR PFAM: PF00036; ehand; 2.
DR PROSITE: PS00018; EF_HAND; 1.
KW Calcium-binding.
FT CA_BIND 412 423 POTENTIAL.
SQ SEQUENCE 467 AA; 48976 MW; 2D2055D1E344FC18 CRC64;

Query Match 14.4%; Score 132.5; DB 1; Length 467;
Best Local Similarity 30.5%; Pred. No. 0.011;
Matches 57; Conservative 9; Mismatches 74; Indels 47; Gaps 12;
```

```
RESULT 3
RBP0_ARATH STANDARD; PRT; 1860 AA.
ID RBP0_ARATH
AC P31635;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6)
DE (VERSION 2).
GN RPB205 OR RPII OR RPB1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RX MEDLINE-91355869; Pubmed-2103447;
RA Dietrich M.A., Prenger J.P., Guilfoyle T.J.;
RT "Analysis of the genes encoding the largest subunit of RNA polymerase
RT II in Arabidopsis and soybean.";
RL Plant Mol. Biol. 15:207-223(1990).
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +
CC RNA(N);
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- PTM: THE TANDEN 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.
CC THE PHOSPHORYLATION ACTIVATES POL2.
CC -1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASE WERE
CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
CC III FOR 5S AND TRNA GENES.
CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X52494; CA36735.1; -.
DR PIR: S11960; JDM02..
DR INTERPRO: IPR000684; -.
DR INTERPRO: IPR000722; -.
DR INTERPRO: IPR002879; -.
DR PFAM: PF01854; RNA_POL_A2; 1.
DR PFAM: PF00623; RNA_POL_A; 1.
DR PROSITE: PS00115; RNA_POL_II_REPEAT; 23.
KW transferase; DNA-directed RNA polymerase; Transcription; Zinc; Repeat;
KW DNA-binding; Nuclear protein; Phosphorylation; Zinc-finger.
FT ZN_FING 66 79
FT ZN_FING 319 390 C2H2-TYPE (POTENTIAL).
FT DOMAIN 806 816 BY SIMILARITY.
FT DOMAIN 1558 1828 ALPHA-AMANTININ BINDING.
SQ SEQUENCE 1860 AA; 207178 MW; BD3FE783FA6977D CRC64;

Query Match 13.5%; Score 124; DB 1; Length 1860;
Best Local Similarity 27.1%; Pred. No. 0.16;
Matches 42; Conservative 17; Mismatches 72; Indels 24; Gaps 6;
```

Query Match	13.48;	Score 123.5;	DB 1;	Length 1841.
Best Local Similarity	27.5%;	Pred. No. 0.17;		
Matches 44;	Conservative 18;	Mismatches 71;	Indels 27;	Gaps 7;
QY	8	PROPTY-PVQP---DGNPVYPQT--LHLRQAPRYTDAPRAVSELYRPSFVHPCGAALVPTM	61	

This SWIFT-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/entire_email_to_license@isb-sib.ch).

EMBL: M27431: AAA28868 1 -


```

RX MEDLINE-96399136; PubMed-8805707;
RA Qiu X., Gulp J.S., Dillella A.G., Hellmig B., Hoog S.S., Janson C.A.,
RA Smith W.W., Abdel-Meguid S.A.;
RT "Unique fold and active site in cytomagalovirus protease.";
RL Nature 383:275-279(1996).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.27 ANGSTROMS) OF PROTEASE.
RX MEDLINE-96399137; PubMed-8805708;
RA Shieh H.-S., Kurumbail R.G., Stevens A.M., Stegeman R.A.,
RA Stutman E.J., Pak J.Y., Wiltner A.J., Palmer M.O., Wiegand R.C.,
RA Hoerwede B.C., Stellingsma W.C.;
RT "Three-dimensional structure of human cytomagalovirus protease.";
RL Nature 383:279-282(1996).
CC -1- FUNCTION: THE CAPSID ASSEMBLY PROTEIN IS A COMPONENT OF THE CAPSID
CC CORE INVOLVED IN PROCESSING AND PACKAGING OF PROGENY DNA.
CC ASSEMBLY IS A PROTEASE WHICH CAN PROTEOLYTICALLY CLEAVE ITSELF
CC AND THE CAPSID ASSEMBLY PROTEIN AT THE C-TERMINUS.
CC -1- CATALYTIC ACTIVITY: PREFERENTIALLY CLEAVES AT ALA-|-SER OR ALA-|-
CC ALA.
CC -1- PHN: CAPSID ASSEMBLY PROTEIN IS PHOSPHORYLATED.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S21 (SERINE PROTEASE).
CC -1- SIMILARITY: TO OTHER HERPESVIRUSES CAPSID PROTEIN VP40.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X17403; CA35353.1; -
DR EMBL: X17403; CA35354.1; ALT-INIT.
DR PIR: S09843; Q08EB8.
DR PDB: 1WPO; 15-OCT-97.
DR PDB: 1LAP; 26-SEP-97.
DR PDB: 1CMV; 04-SEP-97.
DR MEROPS: S21.002; -
DR INTERPRO: IPR001847; -
DR PFAM: PF00716; Nucleo_S21.1.
DR PRINTS: PR00236; HsvCapSIDP40.
DR KW Coat protein; Hydrolyase; Serine protease; Phosphorylation;
KW 3D-structure.
FT CHAIN 1 256 ASSEMBLIN (PROTEASE).
FT CHAIN 257 643 CAPSID ASSEMBLY PROTEIN (BY SIMILARITY).
FT PROP 644 708 C-TERMINAL PEPTIDE (BY SIMILARITY).
FT CHAIN 1 708 GENE UL80 PROTEIN.
FT CHAIN 336 708 GENE UL80.5 PROTEIN.
FT CHAIN 393 708 GENE UL80.4 PROTEIN.
FT CHAIN 478 708 GENE UL80.3 PROTEIN.
FT SITE 256 257 CLEAVAGE (BY THE PROTEASE).
FT SITE 643 644 CLEAVAGE (BY THE PROTEASE).
FT ACT_SITE 63 63 CLEAVAGE (BY THE PROTEASE) (PROBABLE).
FT ACT_SITE 132 132 CHARGE RELAY SYSTEM.
FT ACT_SITE 157 157 CHARGE RELAY SYSTEM.
SQ SEQUENCE 708 AA; 73851 MW; 32A993D568684C9 CAC64;

```

Query Match 12.6%; Score 116.5; DB 1; Length 708;
Best Local Similarity 26.6%; Pred. No. 0.21;
Matches 47; Conservative 17; Mismatches 54; Indels 59; Gaps 9;

```

RESULT 8
RPBL_YEAST
ID RPBL_YEAST STANDARD; PRT; 1733 AA.
AC P04050; Q12364; Q92315;
DT 01-NOV-1986 (Rel. 03, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-1999 (Rel. 38, Last annotation update)
DE DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6) (B220).
GN RPBL OR RPO21 OR RPB220 OR SUV4 OR YDL140C OR D2150.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A364A;
RX MEDLINE-85282617; PubMed-3896517;
RA Allison L.A., Moyle M., Shales M., Ingles C.J.;
RT "Extensive homology among the largest subunits of eukaryotic and
RT prokaryotic RNA polymerases.";
RL Cell 42:599-610(1985).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / FY1679;
RX MEDLINE-97127826; PubMed-8972577;
RA Woelfl S., Haneman V., Saluz H.P.;
RT "Analysis of a 26,756 bp segment from the left arm of yeast
RT chromosome IV.";
RL Yeast 12:1549-1554(1996).
RN [3]
RP SEQUENCE OF 1669-1733 FROM N.A.
RC STRAIN-S288C;
RX MEDLINE-95317607; PubMed-7649444;
RA Cronan J.E. Jr., Wallace J.C.;
RT "The gene encoding the biotin-apolipoprotein ligase of Saccharomyces
RT cerevisiae.";
RL FEMS Microbiol. Lett. 130:221-230(1995).
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE -> N PYROPHOSPHATE +
CC RNA(N).
CC -1- SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.
CC THIS SUBUNIT IS THE LARGEST COMPONENT OF RNA POLYMERASE II.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- PTM: THE TANDDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.
CC THE PHOSPHORYLATION ACTIVATES POL2.
CC -1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE
CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
CC III FOR 5S AND TRNA GENES.
CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X03128; CA26904.1; -
DR EMBL: X96876; CA65619.1; -
DR EMBL: 274188; CA98713.1; -
DR EMBL: U27182; AAC49058.1; -
DR PIR: A00692; RMBY2L.
DR SGD: S0002299; RPO21.
DR INTERPRO: IPR000684; -
DR INTERPRO: IPR000722; -
DR INTERPRO: IPR002879; -
DR PFAM: PF01854; RNA_pol_A2; 1.
DR PFAM: PF00623; RNA_pol_A; 1.

```


CC R Genes Dev. 3:1940-1956(1989).
 CC -1- FUNCTION: ACTS AS A HOMEOTIC SELECTOR GENE CONTROLLING ANTENNAL
 CC AND MANDIBULAR SEGMENT IDENTITY.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- DEVELOPMENTAL STAGE: EMS HAS TWO DIFFERENT SPATIAL PATTERNS OF
 CC EXPRESSION DURING EMBRYOGENESIS. THE EMS HEAD-SPECIFIC EXPRESSION
 CC PATTERN INITIATES PRIOR TO CELLULAR BLASTODERM AND CONTINUES ONLY
 CC UNTIL EARLY GERM-BAND EXTENSION. THE EMS METAMERIC EXPRESSION
 CC PATTERN INITIATES AFTER GASTRULATION AND IS EXPRESSED IN THE
 CC LATERAL NEUROBLASTS, IN ECTODERMAL CELLS AT THE ANTERIOR LATERAL
 CC BORDERS OF EACH SEGMENT, AND IN THE FILIKORPER PRIMORDIA.
 CC -1- MISCELLANEOUS: THE SEQUENCE SHOWN IS THAT OF STRAIN CANTON S.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb.ch/announce/>
 CC or send an email to license@isb.slb.ch).
 CC
 CC EMBL: X66270; CAA46985.1; -
 CC EMBL: X51653; CAA35965.1; -
 CC PIR: A34218; A34218.
 CC PIR: S22708; S22708.
 CC HSSP: P06601; 1FJL.
 CC TRANSFAC: T02008; -
 CC FLYBASE: FBgn0000576; ems.
 CC INTERPRO: IPR001356; -
 CC PFAM: PF00046; homeobox.1.

DR	KM	FT	FT	FT	FT	FT	FT	FT
<hr/>								
DR	PROSITE; PS50711; Homeobox-2, transcription factor.	Homeobox;	DNA-binding; Developmental protein; Nuclear protein.					
FT		1	387					
FT	DOMAIN	99	362					
FT	DOMAIN	334	341					
FT	DNA_BIND	391	450					
FT	DOMAIN	471	497					
FT	VARIANT	4	4					
FT	VARIANT	102	104					
FT	VARIANT	131	131					
FT	VARIANT	138	159					
FT	VARIANT	304	305					
SQ	SEQUENCE	497 AA:	53784 MW;					
			B/E4548155CA5C CR664,					

[illegible]

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=FETAL BRAIN;
 RX MEDLINE=95279403; Pubmed=7759517;
 RA Zhou G., Bao Z.Q., Dixon J.E.;
 RT "Components of a new human protein kinase signal transduction
 pathway";
 RL J. Biol. Chem. 270:12665-12669(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PLACENTA;
 RX MEDLINE=95374539; Pubmed=7646528;
 RA Lee J.-D., Ulevitch R.J., Han J.;
 RT "Primary structure of BMK1, a new mammalian map kinase";
 RL Biochem. Biophys. Res. Commun. 213:715-724(1995).
 CC -1- FUNCTION: MEK5 AND ERK5 INTERACT SPECIFICALLY WITH ONE ANOTHER AND
 NOT WITH MEK1/ERK1 OR MEK2/ERK2 PATHWAYS.
 CC -1- ENZYME REGULATION: ACTIVATED BY TYROSINE AND THREONINE
 PHOSPHORYLATION (BY SIMILARITY).
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN MANY ADULT TISSUES. ABUNDANT IN
 HEART, PLACENTA, LUNG, KIDNEY AND SKELETAL MUSCLE. NOT DETECTABLE
 IN LIVER.
 CC -1- DOMAIN: THE SECOND PROLINE-RICH REGION MAY INTERACT WITH ACTIN
 TARGETING THE KINASE TO A SPECIFIC LOCATION IN THE CELL.
 CC -1- PTM: AUTOPHOSPHORYLATED ON THREONINE AND TYROSINE RESIDUES, WHEN
 THE C-TERMINAL PART OF THE KINASE, WHICH COULD HAVE A REGULATORY
 ROLE, IS ABSENT.
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC MAP KINASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: U25278; AAA81381.1; -
 DR EMBL: U29725; AAA82931.1; -
 DR EMBL: U29726; AAA82932.1; -
 DR EMBL: U29727; AAA82933.1; -
 DR HSSP: P24941; IAO1.
 DR MIM: 602521; -
 DR INTERPRO: IPR000719; -
 DR INTERPRO: IPR002290; -
 DR PFAM: PF00069; PKINASE; 1.
 DR PROSITE: PS01351; MAPK; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00101; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
 KM Transferrase: Serine/chreonine-protein kinase: ATP-binding; Cell cycle;
 KW Phosphorylation.
 FT DOMAIN 54 346
 FT NP_BIND 60 68
 FT BINDING 83 83
 FT ACT_SITE 181 181
 FT MOD_RES 218 218
 FT MOD_RES 220 220
 FT MOD_RES 220 220
 FT DOMAIN 337 340
 FT DOMAIN 433 464
 FT DOMAIN 520 523
 FT DOMAIN 577 699
 FT CONFLICT 19 32
 FT CONFLICT 609 609
 FT SEQUENCE 815 AA; 88636 MW; 379AD69803207CCF CRC64;
 L -> V (IN REF. 2).
 AREGTRPHRLCS -> GPVKVEPAHTASVA (IN
 REF. 2).

Query Match 11.9%; Score 109.5; DB 1; Length 815;
 Best Local Similarity 28.9%; Pred. No. 0.73;
 Matches 52; Conservative 17; Mismatches 56; Indels 55; Gaps 12;
 OY 8 PTPPTVYVPPG---NPVPTLHLRQAPPTTADPRVSLYRPSVHGCAATVPTMSNA 64
 DB 598 PVQPTSP--PPGLAQTGPO---PDSAGSTSGP-----VPQPCPPPGPPAPHPPT--- 642
 OY 65 FPGASLYLPMASVAVGPGSLGSLTPMAVYPVGPITPPGSTVLVGGYDAGARFGAGATAGN 124
 DB 643 GPPGPPIVPAAPQIAT---STSLAAGSLVP--PPG-----LPGSSTPGV 662
 OY 125 IPPPPPGCP---NAQLVAGCANV-LYTKKGN-----FFMG---GSDGQYIT 167
 DB 683 LPYFPPGLPPDAGAPDSSMSSESPVNLVTQLSKSQVEDPLPVFSGHPKSGAGYGV 742

RESULT 15
 MSP8_EIMAC STANDARD; PRT; 259 AA.
 AC P09125;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DE 01-FEB-1994 (Rel. 28, Last annotation update)
 DE MERZOITE SURFACE PROTEIN CMZ-8 (FRAGMENT).
 OS Eimeria acervulina.
 OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriidae; Eimeriidae;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=12;
 RX MEDLINE=89041578; Pubmed=3186453;
 RA Jenkins M.C.;
 RT "A cDNA encoding a merozoite surface protein of the protozoan Eimeria
 acervulina contains tandem-repeated sequences";
 RL Nucleic Acids Res. 16:9863-9863(1988).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: X12445; CAA30977.1; -
 DR PIR: S01704; S01704.
 KW Merzoite; Repeat.
 FT FT 1 1
 FT DOMAIN 5 109
 FT REPEAT 12 18
 FT REPEAT 19 25
 FT REPEAT 26 32
 FT REPEAT 33 39
 FT REPEAT 40 46
 FT REPEAT 47 53
 FT REPEAT 54 60
 FT REPEAT 61 67
 FT REPEAT 68 74
 FT REPEAT 75 81
 FT REPEAT 82 88
 FT REPEAT 89 95
 FT REPEAT 96 102
 FT REPEAT 103 109
 FT SEQUENCE 259 AA; 26988 MW; EFC8B01AACF40502 CRC64;

Query Match 11.8%; Score 108.5; DB 1; Length 259;
 Best Local Similarity 26.2%; Pred. No. 0.29;
 Matches 38; Conservative 18; Mismatches 58; Indels 31; Gaps 6;

```

OY 11 PTYVQPPGNPVYPTLIHPAPPYTDAPPAXSELYRPSF-VHPGATV--PTMSAEPG 67
   | : | | | | | : | : | | | : | : | : | : | : | : | : | : |
Db 7 PSTVSPSPSTPVSPSPS--TPVSPSPSTPVSPSPSTPVSPSPSTPVSPSPSTPVSPSPS 64
OY 68 ASLYLPMAQSVAVGPIGSTIPMAYYPVGPPIYPGSTVLEGGYDAGARFGAGATAGNIPP 127
   | : | | | | | : | : | | | : | : | | | : | : | | | : | : |
Db 65 TPVSPSP--STPVSPSPSTPVSP--PSTPVSPSPSTPV-----SPP 98
OY 128 PPGCGPPNAAQLAVWOGANLVATOR 152
   | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 99 STPVSPSPSPAPAGAVGVNSSLQOR 123

```

Search completed: March 6, 2001, 11:53:31
 Job time: 168 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 11:51:32 ; Search time 47.68 seconds

(Without alignments)
86.058 Million cell updates/sec

Title: US-09-407-430-1_COPY_1_120

Perfect score: 653
Sequence: 1 MNSKGOYPTQPTVQPPGNC.....GSTVLEGGYDAGAFGACA 120

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

1: A.Geneseq_36.*
2: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT.*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	127	19.4	106	21	Y65155
2	101	15.5	1274	20	W69253
3	100	15.3	171	13	R27476
4	99.5	15.2	216	9	P83192
5	99.5	15.2	751	16	R80839
6	98.5	15.1	652	9	P82974
7	98.5	15.1	744	9	P82975
8	98	15.0	246	19	W61955
9	98	15.0	264	19	W71219
10	98	15.0	492	20	V27246
11	96.5	14.8	203	9	P82970
12	96.5	14.8	203	9	P83193

13	96.5	14.8	258	19	W61954
14	96.5	14.8	262	14	R42200
15	95	14.5	650	21	R85180
16	94	14.4	259	17	R97696
17	94	14.4	902	17	W02250
18	94	14.4	1051	20	V29321
19	93.5	14.3	112	21	R84457
20	93.5	14.3	334	9	P82971
21	93.5	14.3	334	9	P83194
22	93.5	14.3	2783	13	R23962
23	93.5	14.3	2783	13	R23963
24	93	14.2	214	17	R86913
25	92	14.1	902	20	W65730
26	91.5	14.0	267	14	R40171
27	90.5	13.9	505	14	R34127
28	90.5	13.9	503	14	R34128
29	90.5	13.9	1181	21	V82707
30	90.5	13.9	1220	21	V82708
31	90	13.8	264	12	R12531
32	90	13.8	278	12	R13338
33	89.5	13.7	137	20	V59678
34	89.5	13.7	137	21	V84606
35	89.5	13.7	572	18	W31855
36	89.5	13.7	763	18	W31852
37	89.5	13.7	1288	18	W26326
38	89.5	13.7	1288	20	W22297
39	89	13.6	155	19	W51548
40	89	13.6	197	21	V66735
41	89	13.6	250	19	W71218
42	88	13.5	259	10	P91655
43	88	13.5	371	19	W64535
44	88	13.5	466	11	R07084
45	88	13.5	626	18	W16398

ALIGNMENTS

Rat galectin amino
Ige binding protei
Cellulose synthase
WB-2 WW domain 11
Human transcriptio
Human btaxin-2 11k
Amino acid sequenc
Bioadhesive precu
Sequence of a b1oa
AFP-1. Homo sapie
AFP-1 (A1A 2460 V
Cotton fiber-speci
Sequence encoded b
Annexin XI type I
Human apoptosis re
Mac2.16 expression
Secreted protein 1
A human small prol
Mycobacterium tube
Mycobacterium tube
Mouse alpha-1 co11
Human secreted pro
Membrane-bound pro
Beta-D-galactoside
Elmeria cell surf
Human leukocyte ce
Recombinant human
Human neuron-deriv

RESULT 1	Y65155	standard; Protein: 106 AA.
XX	Y65155;	
AC	Y65155;	
XX		
DT	01-FEB-2000	(first entry)
XX		
DE	Human 5' EST related polypeptide SEQ ID: NO:1316.	
XX		
KW	Human: 5' EST: expressed sequence tag; secreted protein; diagnosis;	
KW	gene therapy; chromosome mapping; upstream regulatory sequence;	
KW	forensic; location; development; protein synthesis; stability;	
KW	regulation; identification.	
XX		
OS	Homo sapiens.	
XX		
PN	W09953051-A2.	
XX		
PD	21-OCT-1999.	
XX		
PF	09-APR-1999;	99WO-1B00712.
XX		
PR	09-APR-1998;	98US-0057719.
PR	28-APR-1998;	98US-0069047.
XX		
PA	(GENST) GENSET.	
XX		
PI	Dumas Milne Edwards J, Duclert A, Giordano J;	
XX		
DR	WPI: 2000-038446/03.	
DR	N-PSDB: 242769.	
XX		
PT	Novel secreted protein 5' expressed sequence tag sequences used in	
	diagnostic, forensic, gene therapy, and chromosome mapping procedures	

XX Anderson DM, Maugh KJ;
PI WPI: 1992-340250/41.
XX DR N-PSDB; Q29118.
XX Bactericidal coil expression vector - for production of
PT bio-adhesive precursor protein analogues in high yields
XX
PS Example 2; Fig 5; 14pp; English.
XX
CC This sequence represents a recombinantly produced 20 repeat of a
CC bioadhesive decapeptide after cyanogen bromide treatment to
CC separate it from a fusion protein. It is produced by expression of
CC a sequence encoding 20 repeats of the protein (PCX2365) in E. coli
CC GX3015. The protein may be used as a marine, biomedical or dental
CC adhesive. Recombinant production avoids the necessity of handling
CC and processing large quantities of mussels, the normal source of such
CC bioadhesives. The protein is an analogue of non-hydroxylated polyphenolic
CC adhesive proteins produced by *Mytilus edulis*.
XX
SQ Sequence 171 AA:

Query Match 15.3%; Score 100; DB 13; Length 171;
Best Local Similarity 28.8%; Pred. No. 0.028;
Matches 30; Conservative 9; Mismatches 51; Indels 14; Gaps 5;

OY 7 YPQPTYPVQPCGNPVYPTLHL-PCAPPTDAPPAVSELY-----RPSFVHGATVPT 60
DB 64 YKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSY-PPYKAKPS 122
OY 61 MSAAFPGASLYLPMAQSVAVGPIGSTRIPMAVY-----VGPITPP 100
DB 123 YPPYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSYPP 163

RESULT 4
P83192
ID P83192 standard; Protein: 216 AA.
XX
AC P83192;
XX
DT 29-OCT-1992 (first entry)
XX
DE Sequence of the bioadhesive precursor protein analogue encoded by
DE PCX2365, after cyanogen bromide cleavage.
XX
KM Bioadhesive: wound healing; bonding; recombinant adhesive.
XX
OS *Mytilus edulis*.
XX
PN W08807076-A.
XX
PD 22-SEP-1988.
XX
PF 11-MAR-1988; 88WO-US00876.
XX
PR 12-MAR-1987; 87US-0025243.
XX
PA (GENE-) GENEX CORP.
XX
PI Maugh KJ, Anderson DM, Strausberg SL, Strausberg R;
PI Wel T;
XX
DR WPI: 1988-285533/40.
XX
PT Bio-adhesive precursor protein analogues prodn. - by DNA
PT recombinant techniques, subsequent hydroxylation, gives an
PT adhesive for use in wet environment
XX
PS Example; Fig 9; 101pp; English.

CC The bioadhesive precursor protein analogue of the invention can be
CC hydroxylated (prod. claimed) and used as an adhesive in wet
CC environments. The resulting bioadhesives have marine, biomedical and
CC dental applications. They may be used e.g. in wound healing in the
CC same manner as fibrin. While the decapeptide ala-lys-pro-ser-tyr-pro-
CC pro-thr-tyr-lys is repeated many times in the polyphenolic adhesive
CC protein of *M. edulis*, the examination of cDNA sequences encoding
CC portions of this protein (see U.S. Patent Application Serial No. 933,
CC 945) reveals that many other repetitive decapeptide and hexapeptide
CC sequences are also present in the protein, and that these other
CC sequences may constitute the majority of AA sequence of the
CC polyphenolic adhesive protein of *M. edulis*. For example, in cDNA
CC clone 14-1 nineteen decapeptides and one hexapeptide are encoded
CC PCX2354 (10-decapeptide repeats) was used for the construction of
CC PCX2358 (15-decapeptide repeats) and PCX2358 was used for the
CC construction of PCX2365 (20-decapeptide repeats). The amino acid
CC sequence of the 24,077 mol. wt. bioadhesive precursor protein
CC analogue produced by cells containing plasmid PCX2365 after
CC cyanogen bromide cleavage is given in P83192.
XX
SQ Sequence 216 AA:

Query Match 15.2%; Score 99.5; DB 9; Length 216;
Best Local Similarity 28.4%; Pred. No. 0.04;
Matches 31; Conservative 10; Mismatches 49; Indels 19; Gaps 5;

OY 7 YPQPTYPVQPCGNPVYPTLHL-PCAPPTDAPPAVSELY-----RPSFVHGATVPT 60
DB 84 YKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSY-PPYKAKPS 142
OY 61 MSAAFPGASLYLPMAQSVAVGPIGSTRIPMAVY-----VGPITPP 100
DB 143 YPPYKAKPSYPPYKAKPSYPPYKAKPSYPPYKAKPSYPP 188

RESULT 5
R80839
ID R80839 standard; Protein: 751 AA.
XX
AC R80839;
XX
DT 11-JUL-1996 (first entry)
XX
DE Japanese sea mussel adhesive protein.
XX
KM Bioadhesive: Gastropoda.
XX
OS *Mytilus galloprovincialis*.
XX
FH Key Location/Qualifiers
FH Region 101..110
FT /label= repeat
FT /note= "first of many tandem AKPSYPTVK-like repeats"
XX
PN JP07163361-A.
XX
PD 27-JUN-1995.
XX
PF 14-DEC-1993; 93JP-0313338.
XX
PR 14-DEC-1993; 93JP-0313338.
XX
PA (KAIY-) KAIYO BIOTECHNOLOGY KENKYUSHO KK.
XX
XX WPI: 1995-260043/34.
XX
DR N-PSDB; Q99774.
XX
PT Novel gene encoding a Japanese sea mussel adhesive protein - useful
PT as a material for the manufacture of adhesives.
XX
PS Claim 1; Page 6; 6pp; Japanese.

XX	Y27246;
AC	08-OCT-1999 (first entry)
DT	C. elegans CED-6 protein.
XX	
DE	
XX	CED-6; phagocytosis; C. elegans; human; h1CED-6; h2CED-6; h3CED-6;
XX	signal transduction pathway; apoptotic cell; inflammation; cancer;
KW	autoimmune disease; neurodegenerative disease; sickle cell anemia;
KM	Huntington's disease; Alzheimer's disease; myocardial heart failure;
OS	amyotrophic lateral sclerosis; stroke; AIDS.
XX	
XX	Caenorhabditis elegans.
FH	Location/Qualifiers
FT	Key 49..193
FT	/note= "specifically claimed phosphotyrosine binding
FT	(PTB) domain (Y27247)"
FT	Region 194..239
FT	/note= "charged region"
FT	242..338
FT	/note= "specifically claimed proline/serine
FT	rich region (Y27248)"
PN	WO937770-A1.
PD	29-JUL-1999.
XX	
PF	21-JAN-1999; 99WO-US01361.
XX	
PR	24-SEP-1998; 98GB-0020816.
PR	23-JAN-1998; 98US-0082324.
PR	11-JUN-1998; 98GB-0012660.
PR	11-JUN-1998; 98US-0096347.
XX	11-JUN-1998; 98US-0096731.
PA	(COLD-) COLD SPRING HARBOR LAB.
PA	(DEVG-) DEGEN NV.
PX	Bogaert TAOE, Hengartner MO, Liu Q, Van Criekinge MMR;
PX	WIPI; 1999-469130/39.
DR	N-PSTDB; X99301.
XX	
PT	New isolated Caenorhabditis elegans and human CED-6 genes, used to
PT	develop products for treating, e.g. inflammation or autoimmune
PT	disease
PS	Claim la; Fig 2A; 145pp; English.
XX	
CC	The invention provides isolated Caenorhabditis elegans and human CED-6
CC	phagocytosis genes. The nucleic acid sequences encode C. elegans CED-6
CC	polypeptides and human homologues h1CED-6, h2CED-6 and h3CED-6. The CED-6
CC	polypeptides act as adaptor molecules in a signal transduction pathway
CC	that mediates engulfment of apoptotic cells. Transgenic mammalian cells
CC	transfected with the nucleic acids encoding the CED-6 polypeptides can be
CC	used for determining whether a compound is an inhibitor or an enhancer of
CC	a signal transduction pathway which promotes phagocytosis of apoptotic
CC	cells. The enhancers can be used for treating inflammation, autoimmune
CC	disease or cancer. The inhibitors can be used for treating sickle cell
CC	anemia, stroke or neurodegenerative disease. In particular the products
CC	can be used for treating amyotrophic lateral sclerosis, Huntington's
CC	disease, Alzheimer's disease, myocardial heart failure and AIDS. The
CC	products can also be used for detection and diagnosis. The present
CC	sequence represents the amino acid sequence of C. elegans CED-6 protein.
SQ	Sequence 492 AA;
XX	
Query Match	15.0%; Score 98; DB 20; Length 492;
Best Local Similarity	30.8%; Prod. No. 0.14;
Matches 3/7; Conservative	15; Mismatches 56; Indels 12; Gaps 5

QY 2 NSKQCF--QPIYP-VQPPGNVYVQTLHPQAPRYTDAPVPSLVPSFVHG--AA 56
 Db 233 nskadyenttppilypgilppilisp---mpqppnlpssipsmpandlptlmap 288
 QY 57 TVPMASAFPGASVLYLPMASQVAVPLGS---TIPMAVYVPGIYPPGSTVLVEGCTDAG 113
 Db 289 tipqststnsqgapspsvspastspsgpapslpppprpalappvpabrnpvvspxkntag 348
 RESULT 11
 ID P82970 standard; protein; 203 AA.
 XX P82970:
 AC 03-DEC-1990 (first entry)
 DT
 XX
 DE Bioadhesive precursor protein from cDNA clone 14-1.
 XX Bioadhesive precursor protein; cDNA clone 14-1; mussels; barnacles;
 KM oysters; biomedical adhesive; sealants; wound healing.
 XX
 XX Mytilus edulis.
 OS
 PN WO8803953-A.
 PD 02-JUN-1988.
 XX
 XX 24-NOV-1987; 87WO-US03048.
 PF
 XX 07-AUG-1987; 87US-0082456.
 PR
 XX
 XX (GENE-) GENEX CORP.
 PA
 XX Maugh KJ, Anderson DM, Strausberg R, Strausberg SL, McCandliss R,
 PI WPI: 1988-161622/23.
 XX
 DR N-PSDB: N80228.
 XX
 PT DNA coding for bio-adhesive precursor protein - obtd. from mussels,
 XX barnacles and oysters and used esp. to prepare biomedical adhesive.
 XX
 PS Disclosure; ; pp; English.
 XX
 CC mRNA from the phenol gland of the foot of M. edulis is isolated for
 CC the synthesis of cDNA. The cDNA is inserted into a bacteriophage
 CC and E. coli transfectant and cultured. Five clones were isolated and
 CC characterized, one of which is clone 14-1. The sequence has been
 CC directed to show 20 subunits, including 19 decapeptides and 1 hexapeptide.
 CC Clone 14-1 and 52 (N80229) display extensive homology at their 3' ends.
 CC In particular, the last 138 codons of the translated regions are
 CC identical and include codons for a hexapeptide followed by 12
 CC decapeptides (beginning at codon 75 of clone 14-1 and codon 205
 CC of clone 52. The DNA segment contg. the coding regions of
 CC lambda phage precursor protein of clone 14-1 and 52 was obtained from
 CC lambda g10 cloning vectors by EcoRI digestion and subcloned into the
 CC EcoRI site of plasmid pGX2627.
 CC The protein is chemically or enzymatically hydroxylated to produce
 CC adhesive compns. having excellent properties in wet environments. The
 CC product is partic. useful as a biomedical adhesive or sealant e. g. in
 CC wound healing or as a marine or dental adhesive. It can be used as a
 CC thin film membrane.
 CC See also N80228-34.
 CC
 XX
 SQ Sequence 203 AA:
 Query Match 14.8%; Score 96.5; DB 9; Length 203;
 Best Local Similarity 27.0%; Pred. No. 0.071;
 Matches 31; Conservative 8; Mismatches 55; Indels 21; Gaps 4.
 Y YPQPTTYVQPGNPVYVQTLHPQAPRYTDAPVPSLVPSFVHGCAATVPTMSAF 65

Db 31 ykapsypactytkakpsypptytkakpsypptytkakpsypptytkakpsypptytkak 90
 Oy 66 PG-----ASLYLPMAQSVAVGPG-----STIPMAV-----YVGPPIYP 100
 Db 91 psypptykpkpsypptykkskalsypsyykpkklyppptykpkklyppptykpkpsyp 145

RESULT 12

P83193
 ID P83193 standard; Protein; 203 AA.

XX AC P83193;

XX DT 06-MAR-1992 (first entry)

XX DE Sequence of a bioadhesive precursor protein encoded by cDNA clone 14-1.

XX KW Bioadhesive; wound healing; bonding; recombinant adhesive.

XX OS Mytilus edulis.

XX PN W08807076-A.

XX PD 22-SEP-1988.

XX PF 11-MAR-1988; 88WO-US00876.

XX PR 12-MAR-1987; 87US-0025243.

XX PA (GENE-) GENEX CORP.

XX PI Maugh KJ, Anderson DM, Strausberg SL, Strausberg R;
 XX PT Wei T;

XX DR WPI; 1988-285533/40.
 XX DR N-PSDB; N82449.

XX PT Bio-adhesive precursor protein analogues prodn. - by DNA
 XX PT recombinant techniques, subsequent hydroxylation, gives an
 XX PT adhesive for use in wet environment

XX PS Example; Fig 13; 101pp; English.

XX CC The bioadhesive precursor protein analogue of the invention can be
 CC hydroxylated (prod. claimed) and used as an adhesive in wet
 CC environments. The resulting bioadhesives have marine, biomedical and
 CC dental applications. They may be used e.g. in wound healing in the
 CC same manner as fibrin. While the decapeptide ala-lys-pro-ser-tyr-pro-
 CC pro-thr-tyr-lys is repeated many times in the polyphenolic adhesive
 CC protein of M.edulis, the examination of cDNA sequences encoding
 CC portions of this protein (see U.S. Patent Application Serial No. 933,
 CC 945) reveals that many other repetitive decapeptide and hexapeptide
 CC sequences are also present in the protein, and that these other
 CC sequences may constitute the majority of AA sequence of the
 CC polyphenolic adhesive protein of M.edulis. For example, in cDNA
 CC clone 14-1 nineteen decapeptides and one hexapeptide are encoded.
 CC XX

Sequence 203 AA;

Query Match 14.8%; Score 96.5; DB 9; Length 203;
 Best Local Similarity 27.0%; Pred. No. 0.071;

Matches 31; Conservative 8; Mismatches 55; Indels 21; Gaps 4;

Oy 7 YFTQPTPYQPQPGNPPYPTQTLH-PQAPRYTDAPPAVSELYRPSFVHPSAAVPTMSAA 65
 Db 31 ykapsypactytkakpsypptytkakpsypptytkakpsypptytkakpsypptytkak 90
 Oy 66 PG-----ASLYLPMAQSVAVGPG-----STIPMAV-----YVGPPIYP 100
 Db 91 psypptykpkpsypptykkskalsypsyykpkklyppptykpkklyppptykpkpsyp 145

RESULT 13

W61954
 ID W61954 standard; protein; 258 AA.

XX AC W61954;

XX DT 18-SEP-1998 (first entry)

XX DE Rat galectin amino acid sequence.

XX KW Mortalin; galectin; diabetes-mediating protein; insulin; DMP;
 XX KW diabetes; drug screening assay.

XX OS Rattus sp.

XX PN W09820124-A2.

XX PD 14-MAY-1998.

XX PF 24-OCT-1997; 97WO-IB01627.

XX PR 18-JUL-1997; 97US-0897098.

XX PR 25-OCT-1996; 96US-0029324.

XX PR 05-NOV-1996; 96US-0030088.

XX PR 05-NOV-1996; 96US-0030186.

XX PA (ANDE/) ANDERSEN H U.

XX PA (CHRI/) BIERRE CHRISTENSEN U.

XX PA (FEYS/) FEY S J.

XX PA (KARL/) KARLSEN A E.

XX PA (LARS/) MOSE LARSEN P.

XX PA (NERU/) NERUP J.

XX PA (POCI/) POCIOT F.

XX XX

XX PI Andersen HU, Bjerre Christensen U, Fey SJ, Karlsten AE;

XX PI Mose Larsen P, Nerup J, Pociot F;

XX DR WPI; 1998-286940/25.

XX XX

XX PT Identification of diabetes-mediating protein(s) - by transplanting

XX PT insulin-secreting cells into host at risk of developing diabetes and

XX PT analysing protein expression in transplanted cells

XX PS Disclosure; Fig 4; 154pp; English.

XX XX

XX CC This represents the amino acid sequence of murine mortalin. This is a

XX CC diabetes-mediating protective protein used in the method of invention.

XX CC The invention provides methods for in vivo identification of a diabetes-

XX CC mediating protein (DMP) by transplanting insulin-secreting cells into

XX CC host at risk of developing diabetes and analysing protein expression in

XX CC transplanted cells. The DMPs are useful in drug screening assays for

Sequence 258 AA;

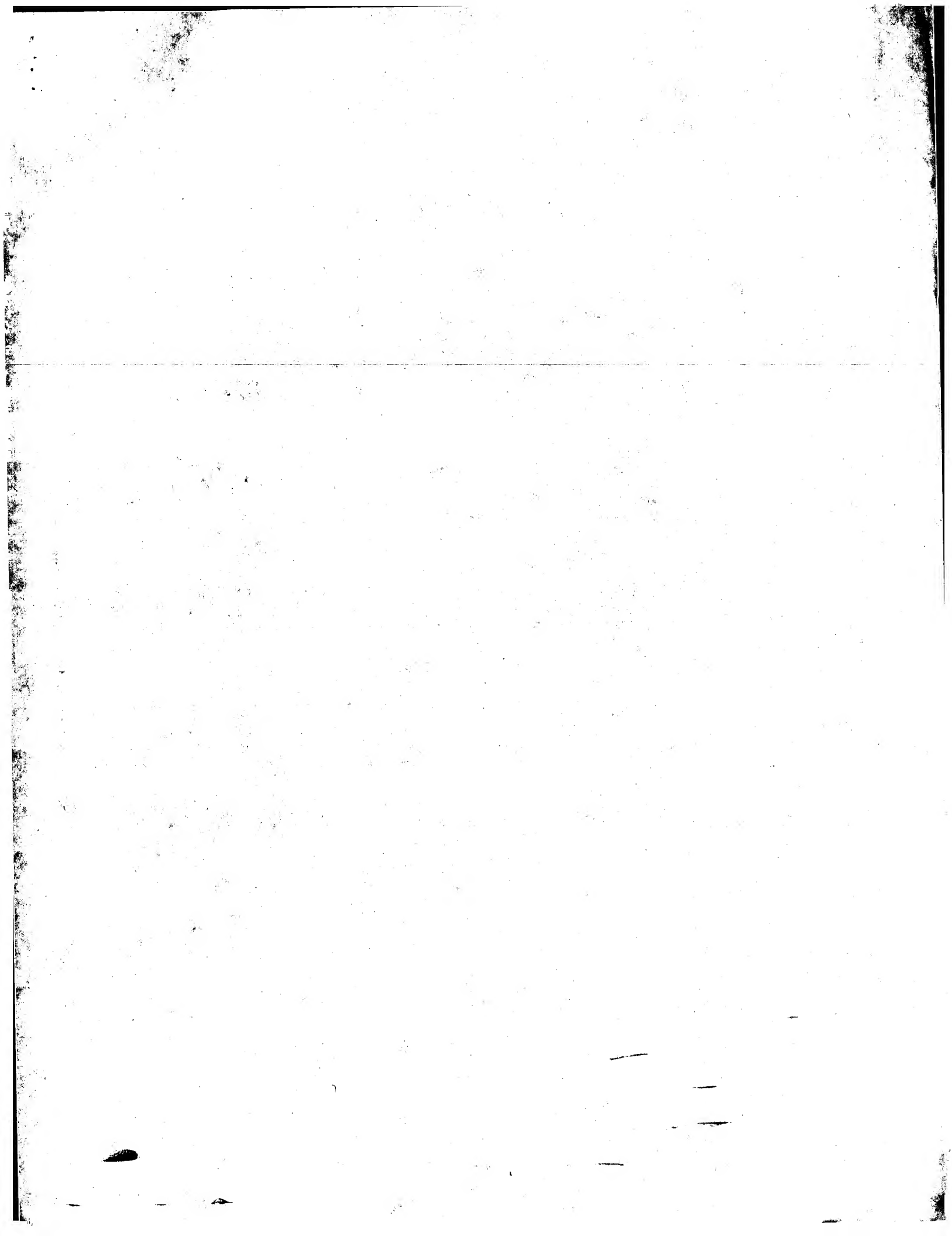
Query Match 14.8%; Score 96.5; DB 19; Length 258;
 Best Local Similarity 33.6%; Pred. No. 0.093;

Matches 40; Conservative 2; Mismatches 40; Indels 37; Gaps 8;

Oy 5 GQYP--TQPTPVQ-PQGNPPYPTQTLH-PQAPRYTDAPPAVSELYRPSFVHPSAAVPTMSAA 61
 Db 33 gypptpvtqptpvqpgnppypptqtlhpgaprytdappavseleypsfvhpagaatvpm 84
 Oy 62 SAAFPAGSLYLPMAQSVAVGPGSTIPMAVYPVGP-IYP-PGSTYLVKGGYDAGARFGA 118
 Db 91 psypptykpkpsypptykkskalsypsyykpkklyppptykpkklyppptykpkpsyp 145

Tue Mar 6 12:02:32 2001

us-09-407-430-1_copy_1_120.rag



Matches	31;	Conservative	8;	Mismatches	55;	Indels	21;	Gaps	4
---------	-----	--------------	----	------------	-----	--------	-----	------	---

NUMBER OF SEQUENCES: 39
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/528,762
FILING DATE: 25-MAY-1990
APPLICATION NUMBER: 82,456
FILING DATE: 07-AUG-1987
APPLICATION NUMBER: 933,945
FILING DATE: 24-NOV-1986
APPLICATION NUMBER: 650,128
FILING DATE: 13-SEP-1984
SEQ ID NO: 37
LENGTH: 331
5202236-37

Query Match 14.3% Score 93.5; DB 5; Length 331;
Best Local Similarity 30.8%; Pred. No. 0.075;
Matches 33; Conservative 6; Mismatches 51; Indels 17; Gaps 5;

QY 3 SKGQYPTQPTTVVQPPG--NPVYPTQLHL-----PQAPPYTDAPPVSELYRPSFVHP 53
DB 175 AKRSP--PTYKAKPTKIKPTPTSTYKAKPTKAKPTKAKPTKAKPTKAKPTKAKPTKAKPTK 232
QY 54 GAATVPTMSAARPGASLYLPMASVAVGCLGSTITPMAYYPVGGIYPP 100
DB 233 -----PTYKSKSTYSPSKKPTKPTKPTKPTKPTKPTKPTKPTKPTKPTKPTKPTKPTKPTK 273

RESULT 12
5202236-3
Patent No. 5202236
APPLICANT: MAUGH, KATHY J.; ANDERSON, DAVID M.; STRAUSBERG,
SUSAN L.; MCCANDLISS, RUSS; WEI, TENA; FILPULA, DAVID
TITLE OF INVENTION: METHOD OF PRODUCING BIOADHESIVE
PROTEIN
NUMBER OF SEQUENCES: 39
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/528,762
FILING DATE: 25-MAY-1990
APPLICATION NUMBER: 82,456
FILING DATE: 07-AUG-1987
APPLICATION NUMBER: 933,945
FILING DATE: 24-NOV-1986
APPLICATION NUMBER: 650,128
FILING DATE: 13-SEP-1984
SEQ ID NO: 3
LENGTH: 334
5202236-3

Query Match 14.3% Score 93.5; DB 5; Length 334;
Best Local Similarity 30.8%; Pred. No. 0.076;
Matches 33; Conservative 6; Mismatches 51; Indels 17; Gaps 5;
QY 3 SKGQYPTQPTTVVQPPG--NPVYPTQLHL-----PQAPPYTDAPPVSELYRPSFVHP 53
DB 178 AKRSP--PTYKAKPTKIKPTPTSTYKAKPTKAKPTKAKPTKAKPTKAKPTKAKPTKAKPTK 235
QY 54 GAATVPTMSAARPGASLYLPMASVAVGCLGSTITPMAYYPVGGIYPP 100
DB 236 -----PTYKSKSTYSPSKKPTKPTKPTKPTKPTKPTKPTKPTKPTKPTKPTKPTKPTKPTK 276

RESULT 13
US-08-217-327-4
Sequence 4, Application US/08217327
Patent No. 5474925
GENERAL INFORMATION:
APPLICANT: John, Malyakal E
APPLICANT: Barton, Kenneth A
TITLE OF INVENTION: Immobilized Proteins in Cotton Fiber
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:

ADDRESSEE: Quarles and Brady
STREET: P.O. Box 2113
CITY: Madison
STATE: WI
COUNTRY: USA
ZIP: 53701-2113
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/217,327
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/812,233
FILING DATE: 19-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J
REGISTRATION NUMBER: 27,386
REFERENCE/DOCKET NUMBER: 1122990831
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-251-5000
TELEFAX: 608-251-9166
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 214 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-217-327-4

Query Match 14.2% Score 93; DB 1; Length 214;
Best Local Similarity 29.6%; Pred. No. 0.051;
Matches 29; Conservative 10; Mismatches 51; Indels 8; Gaps 3;
QY 5 GQPTQPTTVVQPPG--NPVYPTQLHL-----PQAPPYTDAPPVSELYRPSFVHPGATVPTKGA 64
DB 23 GOAPSNP--PTSTPATPTPPAS-----TPPTTQAPPTPTATPTPTPTPTPTPTPTPTPTPTPTPTPT 76
QY 65 PGASLYLPMASVAVGCLGSTITPMAYYPVGGIYPPGS 102
DB 77 PPPVSTPPSPPPATPPASPPATPPASP--PPAT 112

RESULT 14
US-07-706-872-3
Sequence 3, Application US/07706872
Patent No. 5237056
GENERAL INFORMATION:
APPLICANT: Fischbach, Gerald D.
TITLE OF INVENTION: Identification of a Protein Which
PROMOTES THE SYNTHESIS OF ACETYLCHOLINE RECEPTORS AND USES
TITLE OF INVENTION: Therefore
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/706,872
FILING DATE: 19910529
CLASSIFICATION: 514

us-09-407-430-1_copy_1_120.rai

```

1  ATTORNEY/AGENT INFORMATION:
2  NAME: Granahan, Patricia
3  REGISTRATION NUMBER: 32,227
4  REFERENCE/DOCKET NUMBER: H991-01
5  TELECOMMUNICATION INFORMATION:
6  TELEPHONE: (617) 861-6240
7  TELEFAX: (617) 861-9540
8  INFORMATION FOR SEQ ID NO: 3:
9  SEQUENCE CHARACTERISTICS:
10 LENGTH: 267 amino acids
11 TYPE: AMINO ACID
12 TOPOLOGY: linear
13 MOLECULE TYPE: protein
14
15 US-07-706-872-3

```

	Query Match	14.0%;	Score 91.5;	DB 1;	Length 267;	
	Best Local Similarity	24.7%;	Pred. No. 0.092;	Mismatches 44;	Indels 65;	Gaps 7.
	Matches	38;	Conservative	7;		
OY	7 YPQPPVYPOP-----PGNPVYPQTLLHQAAPYTDAF--PAYSELYRPS-----					49
Db	46 YPQPQPIPRPNPCGPHNPGYRHNDGYNHYRQNPFCYPRNNPGYPMGGCGYNSSSGSYHN					105
OY	50 -----FYH-PGAATVPYM-----SAAP					66
		:			:	
Db	106 OKPWKPPTKTFKHVAGAAAAAGAVVGILGGYAMGRVNSGMNYHFDRDEYTRMSENSEARIF					165
OY	67 GASLY----LPMQASVAVGPIGSTIPMAAYYPVGP					96
	: : :					
Db	166 NRYIIRDYISSPVDYFVADC-FNIIVTESISIP					198

RESULT 5
 / Sequence 2, Application US/08526136
 / US-08-526-136-2
 / Patent No. 6107089
 / GENERAL INFORMATION:
 / APPLICANT: Towle, Christine A. et al.
 / TITLE OF INVENTION: ANNEXIN XI
 / NUMBER OF SEQUENCES: 36
 / CORRESPONDENCE ADDRESS:
 / ADDRESSEE: Fish & Richardson
 / STREET: 225 Franklin Street
 / CITY: Boston
 / STATE: Massachusetts
 / COUNTRY: U.S.A.
 / ZIP: 02110-2804
 / COMPUTER READABLE FORM:
 / MEDIUM TYPE: 3.5" Diskette, 1.44 MB
 / COMPUTER: IBM PS/2 Model 50Z or 555X
 / OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
 / SOFTWARE: Wordperfect (Version 5.0)
 / CURRENT APPLICATION DATA:
 / APPLICATION NUMBER: US/08/526,136
 / FILING DATE:
 / CLASSIFICATION: 435
 / PRIOR APPLICATION DATA:
 / APPLICATION NUMBER: US/08/214,036
 / FILING DATE:
 / APPLICATION NUMBER: 07/837,775
 / FILING DATE: February 13, 1992
 / APPLICATION NUMBER: 07/764,465
 / FILING DATE: September 23, 1991
 / ATTORNEY/AGENT INFORMATION:
 / NAME: Clark, Paul T.
 / REGISTRATION NUMBER: 30,152
 / REFERENCE/DOCKET NUMBER: 00786/099001
 / TELECOMMUNICATION INFORMATION:
 / TELEPHONE: (617) 542-5070
 / TELEFAX: (617) 542-8906
 / TELEX: 200154
 / INFORMATION FOR SEQ ID NO: 2:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 503
; TYPE: amino acid
; TOPOLOGY: N/A
;
US-08-526-136-2

```

Query Match	13.9%	Score 90.5;	DB 3;	Length 503;
Best Local Similarity	29.5%	Pred. No. 0.24;		
Matches	38;	Conservative	10;	Mismatches 40;
				Indels 41;
				Gaps 9;
QY	5	GQYPTQ----	PTYPVQRP--	GNEVYRQTLRLQAPRYTDAR-----
DB	90	GQPPAQQPPVSPVSGMXP	PPGNGP----	TSGMSYPRPYCAAPFGQPMILPQQQDPFGVPRG
QY	55	AAATVPMTSAAPFGASL	YLPMAQSVANAGRLTIPMAAYP----	VCPIYPRG-----
DB	146	Q-----PPMT--	YPGQSPVPPGQO-----	VPVSTFGTSGSCTVTPANVSPNQFNGRG
QY	103	TVLVEGGYD	111	
DB	191	TITTDASGFD	199	

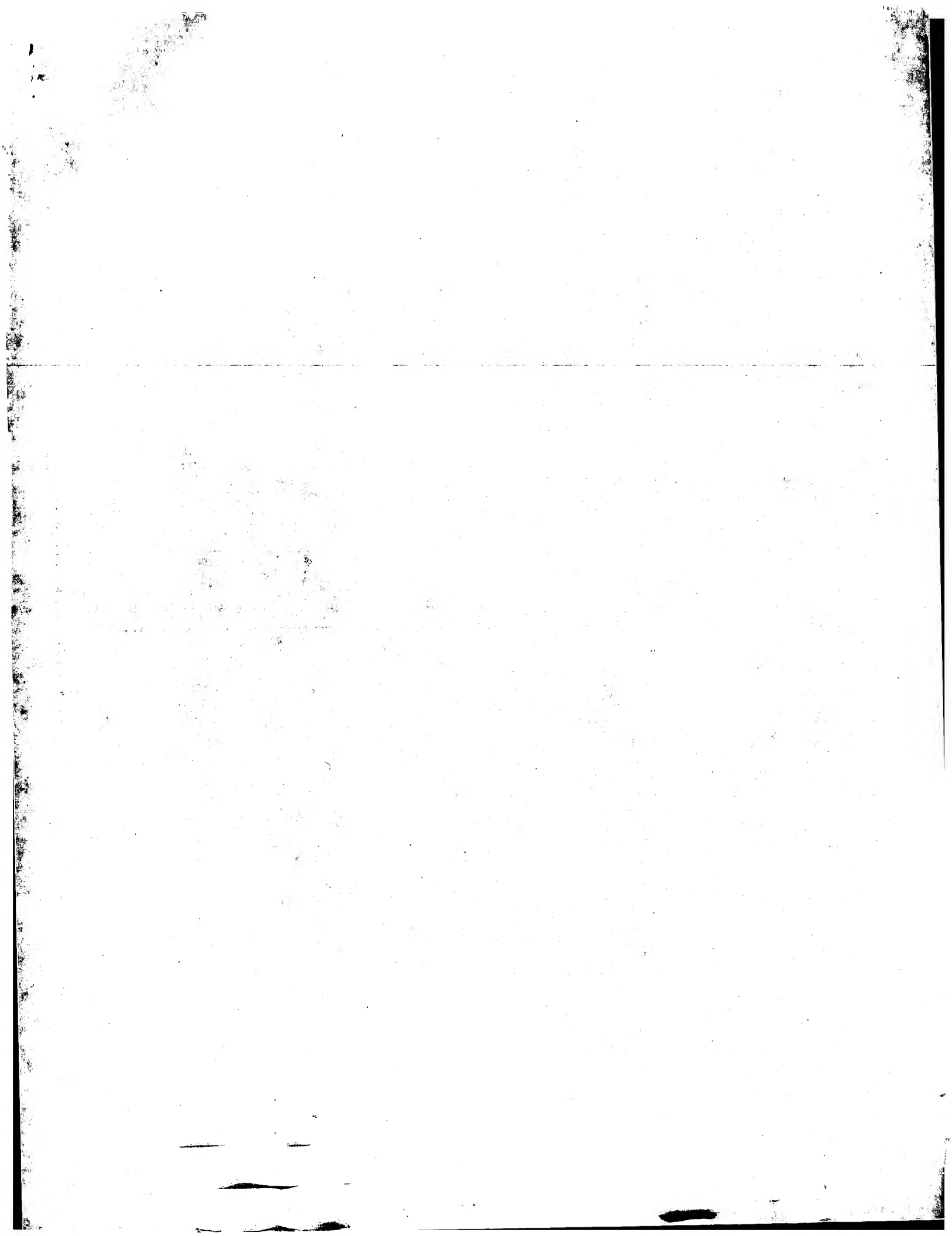
```

Search completed: March 6, 2001, 11:52:12
Job time: 90 sec

```

Tue Mar 6 12:02:32 2001

us-09-407-430-1_copy_1_120.rai



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 11:53:02 ; Search time 45.68 seconds

(without alignments)
178.373 Million cell updates/sec

Title: US-09-407-430-1_COPY_1_120

Sequence: 1 MNSKQYPTQPTVPVPCPN.....GSTVLVEGGYDAGARGACA 120

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 195891.seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR.66:.*
2: PIR1:.*
3: PIR2:.*
4: PIR3:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	121	18.5	491	2	S14182
2	120.5	18.5	625	2	T07796
3	120.5	18.5	650	2	S14181
4	120.5	18.5	977	2	S14183
5	118	18.1	1834	1	JDMU1
6	117.5	18.0	1896	1	JNFE2L
7	116.5	17.8	1553	2	TJ1436
8	116.5	17.8	1733	1	RNBY2L
9	115.5	17.7	1690	2	TJ1670
10	113	17.3	1605	2	TJ1435
11	111.5	17.1	237	2	S46964
12	111.5	17.1	467	2	A27677
13	111.5	17.1	1932	1	A28490
14	111.5	17.1	1970	1	S21054
15	111.5	17.1	1970	2	I38186
16	111	17.0	1859	1	A34092
17	110.5	16.9	502	2	T29959
18	110.5	16.5	620	2	T08776
19	107.5	16.5	708	1	COBEB8
20	107.5	16.5	856	2	TJ1315
21	107.5	16.5	902	2	A56823
22	107.5	16.4	1752	1	S26849
23	105.5	15.9	279	2	T05421
24	104	15.8	2457	2	T18492
25	103.5	15.8	228	2	S4665
26	103	15.7	530	2	T48627
27	102.5	15.7	990	2	T14756
28	101	15.5			

30	101	15.5	2452	1	RN202L	DNA-directed RNA p
31	100.5	15.4	428	2	T24769	hypothetical prote
32	100	15.3	268	2	A49203	homeotic protein C
33	99.5	15.2	92	2	B48831	vitelline membrane
34	99.5	15.2	275	2	T51437	hypothetical prote
35	99.5	15.2	283	2	S13383	hydroxyproline-ric
36	99.5	15.2	751	2	S68957	adhesive plaque pr
37	98	15.0	264	2	A45983	lactose-binding le
38	98	15.0	481	2	T16484	hypothetical prote
39	98	15.0	492	2	T43064	hypothetical prote
40	97.5	14.9	297	2	S23737	proline-rich prote
41	97	14.9	419	2	T29266	hypothetical prote
42	96.5	14.8	262	2	A54889	hypothetical prote
43	96.5	14.8	532	2	T30764	hypothetical prote
44	96.5	14.8	875	2	S23760	hypothetical prote
45	96	14.7	629	2	T05425	polyadenylate-bind

ALIGNMENTS

RESULT 1

S14182

DNA-directed RNA polymerase (EC 2.7.7.6) largest chain (isoform B2) - soybean (fragme

C:Species: Glycine max (soybean)

C>Date: 12-Feb-1993 #sequence-revision 12-Feb-1993 #text-change 12-Sep-1997

R:Dieterich, M.A.; Prenger, J.P.; Gullfoyle, T.J.

Plant Mol. Biol. 15, 207-223, 1990

A:Title: Analysis of the genes encoding the largest subunit of RNA polymerase II in A

A:Reference number: S11960; M01D:91355869

A:Accession: S14182

A:Molecule type: mRNA

A:Residues: 1-491 <DIE>

A:Cross-references: EMBL:X52493

A:Note: translation of the nucleotide sequence is not complete

C:Genetics:

A:Gene: RPB1-B2

C:Superfamily: 416/2; 440/2; 458/2

C:Keywords: DNA binding; nucleic acid transferase; tandem repeat; transcript; zinc f

F:201-474/Region: 7-residue repeats

Query Match

Best Local Similarity 32.7%; Score 121; DB 2; Length 491;

Matches 35; Conservative 11; Mismatches 53; Indels 8; Gaps 4;

Qy 3 SKGQYPTQPTVPVPCPN---NPVYPT--LHLQAPRYTDAPPVSELYRPSFVHGAA 56

Db 239 SPQYSP 297

Qy 57 TTPVMSAATPGASLYLPMAQSA-VGPIGSTRIPMAVYVPGVPGPS 102

Db 298 YSP 344

RESULT 2

T07796

DNA-directed RNA polymerase (EC 2.7.7.6) largest chain - soybean (fragment)

C:Species: Glycine max (soybean)

C>Date: 14-May-1999 #sequence-revision 14-May-1999 #text-change 08-Oct-1999

C:Accession: T07796

R:Dieterich, M.A.; Prenger, J.P.; Gullfoyle, T.J.

Plant Mol. Biol. 15, 207-223, 1990

A:Title: Analysis of the genes encoding the largest subunit of RNA polymerase II in A

A:Reference number: S11960; M01D:91355869

A:Accession: T07796

A>Status: preliminary; translated from CB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-625 <DIE>

A:Cross-references: EMBL:X52492; NID:g18731; PIDN:CA3673.1; PID:g18732

A:Experimental source: strain var. Wayne

A:Gene: SGD:RPO21; MIPS:YDL140C
 A:Cross-references: SGD:S0002299; MIPS:YDL140C
 A:Map position: 4L
 C:Superfamily: human DNA-directed RNA polymerase II largest chain
 C:Keywords: DNA binding; nucleotidyltransferase; tandem repeat; transcription; zinc finger
 F:1567-1733/Region: 7-Residue repeats

Query Match 17.8%; Score 116.5; DB 1; Length 1733;
 Best Local Similarity 31.7%; Pred. No. 0.056;
 Matches 33; Conservative 14; Mismatches 52; Indels 5; Gaps 4;

QY 1 MNSKQYPTQPTV-PVQPGNPNVYPTLHLPOAPPYTDAPPAYSELYRPSFVHGAAATVP 59
 DB 1537 VSSGFGPTSPPTSPPTSPPTSPPTSPPTSPPTSPPTSPPTSPPTSPPTSPPTSPPTSP 1593

QY 60 TMSAAPGASLYLPMQASVA-VGPIGSTITPMAYYPVGPPIPPGS 102
 DB 1594 TSPSYSP 1637

RESULT 9
 T31670
 DNA-directed RNA polymerase (EC 2.7.7.6) II largest chain - Mastigamoeba invertens (frag
 C:Species: Mastigamoeba invertens
 C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
 C:Accession: T31670
 R:Stillier, J.W.; Duffield, E.C.S.; Hall, B.D.
 Proc. Natl. Acad. Sci. U.S.A. 95, 11769-11774, 1998
 A:Title: Antichondriate amoeba and the evolution of DNA-dependent RNA polymerase II.
 A:Reference number: Z21052; MUID:98426228

A:Accession: T31670
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1690 <STI>
 A:Cross-references: EMBL:AF083338; NID:g3511286; PID:g3511287; PIDN:AC62246.1
 A:Experimental source: ATCC50338
 C:Genetics:
 A:Gene: RPB1
 A:Note: Intron positions not resolved (incomplete sequence)
 C:Keywords: DNA binding; nucleotidyltransferase; transcription

Query Match 17.7%; Score 115.5; DB 2; Length 1690;
 Best Local Similarity 31.1%; Pred. No. 0.066; 51; Indels 11; Gaps 4;
 Matches 32; Conservative 9; Mismatches 51;

QY 5 GQYPTQPTV-PVQPGNPNVYPTLHLPOAPPYTDAPPAYSELYRPSFVHGAAATVP 60
 DB 1505 GYSPASPASPASPASPASPASPASPASPASPASPASPASPASPASPASPASPASPASPASP 1558

QY 61 MSAAPGASLYLPMQASVA-VGPIGSTITPMAYYPVGPPIPPGS 102
 DB 1559 SPAYSPASPASPASPASPASPASPASPASPASPASPASPASPASPASPASPASPASPASP 1601

RESULT 10
 T31435
 DNA-directed RNA polymerase (EC 2.7.7.6) II largest chain - Valdimorpha necatrix
 C:Species: Valdimorpha necatrix
 C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-May-2000
 C:Accession: T31435
 R:Hirt, R.P.; Logsdon, J.M.; Healy, B.; Dorey, M.W.; Doolittle, W.F.; Embley, T.M.
 Proc. Natl. Acad. Sci. U.S.A. 96, 580-585, 1999
 A:Title: Microsporidia are related to fungi: evidence from the largest subunit of RNA po
 A:Reference number: Z21034; MUID:99110933
 A:Accession: T31435
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1605 <HIR>
 A:Cross-references: EMBL:AF060234; NID:g4001823; PID:g4001824; PIDN:AD12604.1
 C:Superfamily: human DNA-directed RNA polymerase II largest chain
 C:Keywords: nucleotidyltransferase

Query Match 17.3%; Score 113; DB 2; Length 1605;
 Best Local Similarity 31.4%; Pred. No. 0.1;
 Matches 32; Conservative 12; Mismatches 50; Indels 8; Gaps 4;

QY 8 PQQPTV-PVQPGNPNVYPTLHLPOAPPYTDAPPAYSELYRPSFVHGAAATVP 61
 DB 1476 PISPTSPPTSPPTSPPTSPPTSPPTSPPTSPPTSPPTSPPTSPPTSPPTSPPTSPPTSP 1534

QY 62 SAAAPGASLYLPMQASVA-VGPIGSTITPMAYYPVGPPIPPGS 102
 DB 1535 PSYSP 1576

RESULT 11
 S46964
 Microfilament sheath protein - nematode (Brugia malayi)
 C:Species: Brugia malayi
 C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 17-Mar-2000
 C:Accession: S46964
 R:Hitzmann, J.; Schnauffer, A.; Hintz, M.; Contraths, F.J.; Stilm, S.; Zahner, H.; Hob
 submitted to the EMBL Data Library, July 1994
 A:Description: Brugia spp. and Litomosoides carinii: characterization of the shp2 gen
 A:Reference number: S46964
 A:Accession: S46964
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-237 <HIR>
 A:Cross-references: EMBL:Z25444; NID:g516352; PID:g516353
 C:Genetics:
 A:Introns: 26/1
 C:Superfamily: hydroxyproline-rich glycoprotein

Query Match 17.1%; Score 111.5; DB 2; Length 237;
 Best Local Similarity 35.6%; Pred. No. 0.018; 39; Indels 11; Gaps 5;
 Matches 32; Conservative 8; Mismatches 39;

QY 6 QYPTQPTV-PVQPGNPNVYPTLHLPOAPPYTDAPPAYSELYRPSFVHGAAATVP 58
 DB 128 QYPTQPT 185

QY 59 PT-MSAAPGASLYLPMQASVAVGPIGSTIT 87
 DB 186 PQYPTQPT 214

RESULT 12
 A27677
 DNA-directed RNA polymerase (EC 2.7.7.6) II largest chain - Chinese hamster (fragment
 C:Species: Cricetus griseus (Chinese hamster)
 C:Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 18-Jun-1999
 C:Accession: A27677
 R:Allison, L.A.; Wong, J.K.C.; Fitzpatrick, V.D.; Moyle, M.; Ingles, C.J.
 Mol. Cell. Biol. 8, 321-329, 1988
 A:Title: The C-terminal domain of the largest subunit of RNA polymerase II of Sacchar
 A:Reference number: A93104; MUID:88094402
 A:Accession: A27677
 A:Molecule type: DNA
 A:Residues: 1-467 <ALL>
 A:Cross-references: GB:M9538; NID:g191170; PIDN:AAA37008.1; PID:g387055
 C:Genetics:
 A:Gene: RPO21
 A:Introns: 33/1; 83/1
 C:Superfamily: human DNA-directed RNA polymerase II largest chain
 C:Keywords: DNA binding; nucleotidyltransferase; tandem repeat; transcription; zinc f

Query Match 17.1%; Score 111.5; DB 2; Length 467;
 Best Local Similarity 31.6%; Pred. No. 0.036; 48; Indels 5; Gaps 4;
 Matches 30; Conservative 12; Mismatches 48;

Tue Mar 6 12:02:32 2001

us-09-407-430-1_copy_1_120.rpt

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 11:53:31 ; Search time 25.34 Seconds
(Without alignments)
152.932 Million cell updates/sec

Title: US-09-407-430-1_COPY_1_120
Perfect score: 653
Sequence: 1 MNSKQVPTQPTVQVPPGN.....GSTVLVEGGYDAGARFGAGA 120

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues
Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_39.*
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	118	18.1	1841	RPB1_ARATH	P18616 arabidopsis
2	118	18.1	1860	RPB0_ARATH	P31635 arabidopsis
3	117.5	18.0	1896	RPB1_DROME	P31635 drosophila
4	117	17.9	467	CBPA_DICDI	P35085 dictyostell
5	116.5	17.8	1733	RPB1_YEAST	P04050 saccharomyc
6	116.5	17.1	467	RPB1_CRIGR	P14144 cricetulus
7	111.5	17.1	1970	RPB1_HUMAN	P24328 homo sapien
8	111.5	17.1	1970	RPB1_MOUSE	P08775 mus musculu
9	111	17.0	1859	RPB1_CAEEL	P16356 caenorhabd1
10	107.5	16.5	620	EXTN_TOBAC	P13983 nicotiana t
11	107.5	16.5	708	VP40_HCMVA	P16753 human cytom
12	107.5	16.5	902	RPB1_DICDI	P35084 dictyostell
13	107	16.4	1752	RPB1_SCHPO	P36594 schizosacch
14	104	15.9	346	PRE1_LYCES	Q00451 lycopersico
15	101	15.5	2452	RPB1_PLAFD	P14248 plasmodium
16	100.5	15.4	872	FPL_MYTGO	Q25434 mytilus cor
17	100	15.3	268	CDX1_MOUSE	P18111 mus musculu
18	99.5	15.2	283	EXTN_SORBI	P24152 sorghum bic
19	99.5	15.2	751	FPL_MYTGA	Q27409 mytilus gal
20	96.5	14.8	261	LEG3_RAT	P08699 rattus norv
21	96.5	14.8	875	FPL_MYTED	Q25460 mytilus edu
22	96	14.7	629	PAB2_ARATH	P42731 arabidopsis
23	94.5	14.5	902	IF2_MYTCTU	P71613 mycobacteri
24	94	14.4	902	NFCA_HUMAN	P14934 homo sapien
25	93	14.2	295	LEG3_CANFA	P34886 canis famli
26	93	14.2	503	ANYA_MOUSE	P33477 oryctolagus
27	92	14.1	541	TELL_DROME	P09957 drosophila
28	91	13.9	503	ANYA_RABIT	P97384 mus musculu
29	91	13.9	1048	SR44_RAT	Q65627 rattus norv
30	90.5	13.9	503	ANXA_BOVIN	P27214 bos taurus
31	90.5	13.9	505	ANXB_BOVIN	P27215 bos taurus
32	90.5	13.9	1183	DREL_RAT	P52258 rattus norv
33	90.5	13.9	2142	BAT2_HUMAN	P48634 homo sapien

ALIGNMENTS

RESULT . 1	ID	STANDARD:	PRT:	1841 AA.
RPB1_ARATH	RPB1_ARATH			
AC	P18616;			
DT	01-NOV-1990 (Rel. 16, Created)			
DT	01-NOV-1990 (Rel. 16, Last sequence update)			
DT	15-DEC-1998 (Rel. 37, Last annotation update)			
DE	DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6)			
DE	(VERSION 1)			
GN	RPB205 OR RPB1 OR RPB1.			
OS	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;			
OC	Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;			
OC	Brassicales; Brassicaceae; Arabidopsis.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRATIN-CV. COLUMBIA;			
RX	MEDLINE-9108067; PubMed-2259344;			
RA	Nawrath C., Schell J., Koncz C.;			
RT	"Homologous domains of the largest subunit of eucaryotic RNA			
RT	polymerase II are conserved in plants.";			
RL	Mol. Gen. Genet. 223:65-75(1990).			
CC	-1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION			
CC	OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS			
CC	SUBSTRATES.			
CC	-1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE -> N PYROPHOSPHATE +			
CC	RNA(N).			
CC	-1- SUBCELLULAR LOCATION: NUCLEAR.			
CC	-1- PTM: THE TANDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.			
CC	THE PHOSPHORYLATION ACTIVATES POL2.			
CC	-1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES WERE			
CC	FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA			
CC	PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE			
CC	III FOR 5S AND TRNA GENES.			
CC	-1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.			
CC	THIS SWISS-PROT ENTRY IS COPYRIGHTED. IT IS PRODUCED THROUGH A COLLABORATION			
CC	BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION			
CC	AT THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS			
CC	USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY			
CC	MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL			
CC	ENTITIES REQUIRES A LICENSE AGREEMENT (SEE http://www.isb-sib.ch/announce/			
CC	OR SEND AN EMAIL TO license@sib-sib.ch).			
CC	-----			
DR	EMBL: X52954; CAA37130.1; -			
DR	PIR: S12071; JDM01.			
DR	INTERPRO: IPR000684; -			
DR	INTERPRO: IPR000722; -			
DR	INTERPRO: IPR002879; -			
DR	PFAM: PF01854; RNA_POL_A2; 1.			
DR	PFAM: PF00623; RNA_POL_A; 1.			
DR	PROSITE: PS00115; RNA_POL_II_REPEAT; 24.			
KW	Transferase; DNA-directed RNA polymerase; Transcription; Zinc; Repeat;			
KW	DNA-binding; Nuclear protein; Phosphorylation; Zinc-finger.			
FT	ZN_FING 66 79			
FT	DNA_BIND 327 398			
FT	BY SIMILARITY.			

34	90	13.8	263	1	LEG3_MOUSE	P16110 mus musculu
35	89.5	13.7	1527	1	CAH1_MOUSE	P39061 mus musculu
36	88	13.6	249	1	LEG3_HUMAN	P17931 homo sapien
37	89	13.6	497	1	HME5_DROME	P18488 drosophila
38	89	13.6	512	1	GAG_SMSAV	P03330 simlen sarc
39	88.5	13.6	273	1	PRIO_CHICK	P27177 gallus gall
40	88.5	13.6	309	1	HXA4_CHICK	P17277 gallus gall
41	88	13.5	259	1	MSR8_ETMAC	P09125 elmeria ace
42	88	13.5	265	1	CDX1_HUMAN	P47902 homo sapien
43	88	13.5	466	1	ANX7_HUMAN	P20073 homo sapien
44	88	13.5	1157	1	SR44_HUMAN	O95104 homo sapien
45	87.5	13.4	435	1	BRAC_HUMAN	O15178 homo sapien

DT 01-NOV-1986 (Rel. 03, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DN DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6) (B220).
GN RPB1 OR RPO21 OR RPB220 OR SUV4 OR YDL140C OR D2150.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-A364A;
RX MEDLINE=85282617; PubMed=3896517;
RA Allison L.A., Moyle M., Shales M., Ingles C.J.;
RT "Extensive homology among the largest subunits of eukaryotic and
RL prokaryotic RNA polymerases.";
Cell 42:599-610(1985).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-S28BC / FYJ679;
RX MEDLINE=97127826; PubMed=8972577;
RA Moelfel S., Haneman V., Saluz H.P.;
RT "Analysis of a 26,756 bp segment from the left arm of yeast
RL chromosome IV.";
yeast 12:1549-1554(1996).
[3]
RP SEQUENCE OF 1669-1733 FROM N.A.
RC STRAIN-S28BC;
RX MEDLINE=95377607; PubMed=7649444;
RA Cronan J.E. Jr., Wallace J.C.;
RT "The gene encoding the biotin-apoprotein ligase of Saccharomyces
RL cerevisiae.";
FEMS Microbiol. Lett. 130:221-230(1995).
-1 FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
SUBSTRATES.
-1 CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE - N PYROPHOSPHATE +
RNA(N).
-1 SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.
THIS SUBUNIT IS THE LARGEST COMPONENT OF RNA POLYMERASE II.
-1 SUBCELLULAR LOCATION: NUCLEAR.
-1 PTM: THE TANDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.
THE PHOSPHORYLATION ACTIVATES POLD2.
-1 MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE
FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA


```

RESULT 9
RPBL CAEEL STANDARD; PRT: 1859 AA.
ID RPBL CAEEL
AC P16356;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6).
GN AMA-1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2; PubMed=2586513;
RA MEDLINE=90066416;
RA Bird D.M., Riddle D.L.;
RT "Molecular cloning and sequencing of ama-1, the gene encoding the
RT largest subunit of Caenorhabditis elegans RNA polymerase II.";
RL Mol. Cell. Biol. 9:4119-4130(1989).
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +
CC RNA(N).
CC -1- SUBUNIT: EACH CLASS OF RNA POLYMERASE IS ASSEMBLED FROM NINE TO
CC FOURTEEN DIFFERENT POLYPEPTIDES.
CC -1- PTM: THE TANDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.
CC THE PHOSPHORYLATION ACTIVATES POL2.
CC -1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES WERE
CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
CC III FOR 5S AND 5.8S GENES.
CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M29235; AAA28126.1;
CC PIR: A34092; A34092.
CC INTERPRO: IPR000684;
CC INTERPRO: IPR000722;
CC INTERPRO: IPR002879;
CC PIRAM: PF01854; RNA_POL_A2.1.
CC PRAM: PF00623; RNA_POL_A1.1; REPEAT: 26.
CC PROSITE: PS00115; RNA_POL_II_REPEAT.
CC TRANSFERASE: DNA-directed RNA polymerase; transcription; zinc; Repeat;
CC DNA-binding; Nuclear protein; Phosphorylation; zinc-finger.
CC ZN_FING 66 82 C2H2-TYPE (POTENTIAL).
CC DOMAIN 1567 1859 CARBOXYL-TERMINAL 7-RESIDUE REPEATS.
CC SEQUENCE 1859 AA; 204631 MW; 756F687B0AACCF8C CRC64;
SQ
Query Match 17.0%; Score 111; DB 1; Length 1859;
Best Local Similarity 29.5%; Pred. No. 0.26;
Matches 33; Conservative 14; Mismatches 57; Indels 8; gaps 5;

```

```

EXTN_TOBAC STANDARD; PRT: 620 AA.
ID EXTN_TOBAC
AC P13983;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-MAR-1992 (Rel. 21, Last annotation update)
DE EXTENSIN PRECURSOR (CELL WALL HYDROXYPROLINE-RICH GLYCOPROTEIN).
GN HRPPT3.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; eusterids I;
OC Solanales; Solanaceae; Nicotiana.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. XANTHI; TISSUE-LEAF;
RC MEDLINE=90128263; PubMed=2612909;
RA Keller B., Lamb C.J.;
RT "Specific expression of a novel cell wall hydroxyproline-rich
RT glycoprotein gene in lateral root initiation.";
RL Genes Dev. 3:1639-1646(1989).
CC -1- FUNCTION: HAS A SPECIALIZED STRUCTURAL FUNCTION, POSSIBLY IN
CC THE MECHANICAL PENETRATION OF THE CORTEX AND EPIDERMIS OF THE
CC MAIN ROOT.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.
CC -1- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE
CC SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN
CC GLYCOSYLATED.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X13885; CAA32090.1;
CC PIR: S06733; S06733.
CC Repeat; Cell wall; Glycoprotein; Signal; Structural protein;
CC HYDROXYLATION.
CC SIGNAL 1 2 3
CC CHAIN 1 2 3
CC REPEAT 70 73 EXTENSIN.
CC REPEAT 148 151 H-A-P-P.
CC DOMAIN 229 242 H-A-P-P.
CC REPEAT 229 235 2 X 7 AA TANDEM REPEATS OF T-H-R-H-A-P-P.
CC REPEAT 236 242 2.
CC DOMAIN 205 620 CONTAINS THE SER-PRO(4) REPEATS.
CC FT 499 600 3 X APPROXIMATE TANDEM REPEATS.
CC DOMAIN 499 600
CC SEQUENCE 620 AA; 65406 MW; 641DD2278AB28524 CRC64;
SQ
Query Match 16.5%; Score 107.5; DB 1; Length 620;
Best Local Similarity 28.9%; Pred. No. 0.16;
Matches 33; Conservative 15; Mismatches 49; Indels 17; gaps 4;

```

RESULT 10

Tue Mar 6 12:02:33 2001

us-09-407-430-1_copy_1_120.rsp

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd

OM protein - protein search, using sw model

```

run on:      march 6, 2001, 11:54:50 ; Search time 75.33 seconds
              (without alignments)
              186.711 million cell updates/sec

```

```

Title:      US-09-407-430-1_COPY_1_120
Perfect score: 653
Sequence:   1 MNSKGOYPTQPTYPVQPPGN.....GSTVLVEGGYDAGAREGAGA 120

```

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing:	Minimum Match	0%
	Maximum Match	100%
	Listing first	45 summaries

```

1:  sp.archaea:*
2:  sp.bacteria:*
3:  sp.fungi:*
4:  sp.inuman:*
5:  sp.invertebrate:*
6:  sp.mammal:*
7:  sp.mhc:*
8:  sp.organelle:*
9:  sp.phage:*
10: sp.plant:*
11: sp.rodent:*
12: sp.virus:*
13: sp.invertebrate:*
14: sp.unclassified:*

```

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	653	100.0	168	4	Q15038	Q15038 homo sapien
2	648	99.2	168	11	O88675	O88675 mus musculus
3	121	18.5	494	10	O99367	O99367 glycine max
4	120.5	18.5	625	10	O99366	O99366 glycine max
5	120.5	18.5	977	10	O99368	O99368 glycine max
6	118	18.1	1840	10	O95258	O95258 arabidopsis
7	117.5	18.0	1811	5	O9XZ08	O9XZ08 drosophila
8	117.5	18.0	1887	5	O9VYX6	O9VYX6 drosophila
9	117.5	18.0	1889	5	O9XZ52	O9XZ52 drosophila
10	117.5	18.0	1889	5	O9XZ51	O9XZ51 drosophila
11	117.5	18.0	1889	5	O9XZ09	O9XZ09 drosophila
12	116.5	17.8	1553	5	O96452	O96452 nosena locu
13	115.5	17.7	1690	5	O77165	O77165 mastigamoeb
14	113.5	17.4	414	3	O94231	O94231 kluweromyc
15	113	17.3	1605	5	O96446	O96446 valtrimorpha
16	111.5	17.1	237	5	O17242	O17242 brugia mala
17	111.5	17.1	1966	11	O08847	O08847 mus musculu
18	111.5	17.1	1970	11	O35559	O35559 cricetus
19	111	17.0	1862	5	Q20090	Q20090 caenorhabdi

[illegible]

ALIGNMENTS

RESULT	1	
015038		
ID	015038	
AC	015038;	PRELIMINARY;
DT	01-NOV-1996 (TEMBLrel. 01, Created)	PRT; 168 AA.
DT	01-NOV-1996 (TEMBLrel. 01, Last sequence update)	
DT	01-JUN-2000 (TEMBLrel. 14, Last annotation update)	
DE	K1AA0058 PROTEIN.	
GN	K1AA0058.	
OS	Homo sapiens (Human).	
CC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
CC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=96051398; PubMed=7584044;	
RA	Monura N., Nagase T., Miyajima N., Sazuka T., Tanaka A., Sato S.,	
RA	Seki N., Kawabayashi Y., Ishikawa K., Tabata S.;	
RT	"Prediction of the coding sequences of unidentified human genes. II.	
RT	The coding sequences of 40 new genes (K1AA0041-K1AA0080) deduced by	
RL	analysis of cDNA clones from human cell line KG-1."	
RL	DNA Res. 1:223-229(1994).	
DR	EMBL; D31767; BAA0545.1; -.	
DR	INTERPRO; IPR000515; -.	
DR	PROSITE; PS00402; BPD_TRANSF_INN_MEMBR; UNKNOWN_1.	
SO	SEQUENCE 168 AA; 17319 MW; 49F1B5D281E24AAC CRC64;	

Query Match	100.0%;	Score 653;	DB 4;	Length 168;
Best Local Similarity	100.0%;	Pred. No. 8.5e-51;		
Matches 120;	Conservative 0;	Mismatches 0;	Indels 0;	Cross 0;

QY	1	2
QY	1 MNSKGGYPTPIQPTPVQPPGPNPNVYPOOTLHLQAPPPYDAPPAISELYRSPFVHPGCAATVPT	60
Db	1 MNSKGGYPTPIQPTPVQPPGPNPNVYPOOTLHLQAPPPYDAPPAISELYRSPFVHPGCAATVPT	60
QY	61 MSAAPFGASLYLPMAQSAVAVPLGSTITPMATYYPVGPITYPGSVTLVEGGDAGARFCGGA	120
Db	61 MSAAPFGASLYLPMAQSAVAVPLGSTITPMATYYPVGPITYPGSVTLVEGGDAGARFCGGA	120

088675 PRELIMINARY; PRT; 168 AA.
 AC 088675;
 DT 01-NOV-1998 (TREMblrel. 08, Created)
 DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
 DE PROLINE-RICH PROTEIN.
 GN PRP OR BPR.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CD1; TISSUE-BRAIN;
 RA Yang W., Mansour S.L.;
 RT "A proline-rich protein expressed in mouse brain";
 RL Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF085348; AAC34594.1; -;
 DR MGD; MGI:1344344; Ptbl.
 DR INTERPRO: IPR000515; -;
 DR PROSITE: PS00402; BPD_TRANSF_INN_MEMBER; UNKNOWN_1.
 DR SEQUENCE 168 AA; 17288 MW; 49EDDAA29D8E344C CRC64;
 SQ

Query Match 99.2%; Score 648; DB 11; Length 168;
 Best local Similarity 99.2%; Pred. No. 2.4e-50;
 Matches 119; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MNSKGQYPTQPTYPVQPPNPVQPTLHLPOAPPTDAPPAVSELYRPSFVHGAATVPT 60
 DB 1 MNSKGQYPTQPTYPVQPPNPVQPTLHLPOAPPTDAPPAVSELYRPSFVHGAATVPT 60
 QY 61 MSAAPGASLYLPMASQVAVGPGSTIPMAYPVPGPIYVPGSTIVGEGYDAGARFGAGA 120
 DB 61 MSAAPGASLYLPMASQVAVGPGSTIPMAYPVPGPIYVPGSTIVGEGYDAGARFGAGA 120

RESULT 3
 099367 PRELIMINARY; PRT; 494 AA.
 ID 099367;
 AC 099367;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
 DE DNA-DIRECTED RNA POLYMERASE II (FRAGMENT).
 GN RPBI-B2.
 OS Glycine max (Soybean).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
 OC Fabales; Fabaceae; Papilionoideae; Glycine.
 OX NCBI_Taxid=3847;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. WAYNE;
 RA Dietrich M.A.;
 RL Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 204-494 FROM N.A.
 RC STRAIN-CV. WAYNE;
 RX MEDLINE=91355869; PubMed=2103447;
 RA Dietrich M.A., Prenger J., Guilfoyle T.J.;
 RT "Analysis of the genes encoding the largest subunit of RNA polymerase
 II in Arabidopsis and soybean";
 RL Plant Mol. Biol. 15:207-223(1990).
 DR EMBL; X52493; CAA36734.1; -;
 DR MENDEL; 16056; Glyma; 238; 16056.
 DR INTERPRO: IPR000684; -;
 DR INTERPRO: IPR002879; -;
 DR INTERPRO: IPR002965; -;
 DR PFAM; PF01854; RNA_POL_A2; 1.
 DR PRINTS; PRO1217; PRICHEXTENS.
 DR PROSITE; PS00115; RNA_POL_II_REPEAT; UNKNOWN_22.
 DR SEQUENCE 494 AA; 52156 MW; 41EEFD030F61D649 CRC64;

KW DNA-directed RNA polymerase.
 FT NON_TER 1
 SQ SEQUENCE 494 AA; 52156 MW; 41EEFD030F61D649 CRC64;

Query Match 18.5%; Score 121; DB 10; Length 494;
 Best local Similarity 32.7%; Pred. No. 0.0033;
 Matches 35; Conservative 11; Mismatches 53; Indels 8; Gaps 4;

QY 3 SKGQYPTQPTYPVQPPG---NPVYPT-----LHLPOAPPTDAPPAVSELYRPSFVHGA 56
 DB 242 SPQYSP 300
 QY 57 TVPTMSAAPGASLYLPMASQVAVGPGSTIPMAYPVPGPIYVPGSTIVGEGYDAGARFGAGA 102
 DB 301 YSPS 347

RESULT 4
 099366 PRELIMINARY; PRT; 625 AA.
 ID 099366;
 AC 099366;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
 DE DNA-DIRECTED RNA POLYMERASE (FRAGMENT).
 GN RPBI-B1 OR GENE B1.
 OS Glycine max (Soybean).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
 OC Fabales; Fabaceae; Papilionoideae; Glycine.
 OX NCBI_Taxid=3847;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. WAYNE;
 RA Dietrich M.A.;
 RL Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 364-625 FROM N.A.
 RC STRAIN-CV. WAYNE;
 RX MEDLINE=91355869; PubMed=2103447;
 RA Dietrich M.A., Prenger J., Guilfoyle T.J.;
 RT "Analysis of the genes encoding the largest subunit of RNA polymerase
 II in Arabidopsis and soybean";
 RL Plant Mol. Biol. 15:207-223(1990).
 DR EMBL; X52492; CAA36733.1; -;
 DR MENDEL; 16055; Glyma; 238; 16055.
 DR INTERPRO: IPR000684; -;
 DR INTERPRO: IPR002879; -;
 DR INTERPRO: IPR002965; -;
 DR PFAM; PF01854; RNA_POL_A2; 1.
 DR PRINTS; PRO1217; PRICHEXTENS.
 DR PROSITE; PS00115; RNA_POL_II_REPEAT; UNKNOWN_23.
 KW DNA-directed RNA polymerase.
 FT NON_TER 1
 SQ SEQUENCE 625 AA; 68251 MW; DC84628BD0F87AF5 CRC64;

Query Match 18.5%; Score 120.5; DB 10; Length 625;
 Best local Similarity 31.6%; Pred. No. 0.0046;
 Matches 36; Conservative 10; Mismatches 53; Indels 15; Gaps 4;

QY 3 SKGQYPTQPTYPVQPPG---NPVYPT-----LHLPOAPPTDAPPAVSELYRPS 49
 DB 395 SPQYSP 453
 QY 50 FVHPGATVPTMSAAPGASLYLPMASQVAVGPGSTIPMAYPVPGPIYVPGSTIVGEGYDAGARFGAGA 102
 DB 454 YSPS 507

RESULT 5
 099368

```

ID 099368 PRELIMINARY: PRT: 977 AA.
AC 099368;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE DNA-DIRECTED RNA POLYMERASE II (FRAGMENT).
GN RPB1-C.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Fabales; Fabaceae; Papilionoideae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. WAYNE;
RA Dieckhoff M.A.;
RL Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 725-977 FROM N.A.
RC STRAIN-CV. WAYNE;
RA MEDLINE-91355869; PubMed-2103447;
RA Dieckhoff M.A., Prenger J., Gullfoyle T.J.;
RT "Analysis of the genes encoding the largest subunit of RNA polymerase
RT II in Arabidopsis and soybean."
RL Plant Mol. Biol. 15:207-223(1990).
DR EMBL: X52495; CAA36736.1;
DR MENDEL: 16057; Glyma:238.16057.
DR INTERPRO: IPR000684;
DR INTERPRO: IPR000722;
DR INTERPRO: IPR002879;
DR PFAM: PF00623; RNA_POL_A: 1;
DR PFAM: PF01854; RNA_POL_A2: 1;
DR PROSITE: PS00115; RNA_POL_II_REPEAT; UNKNOWN_17.
KW DNA-directed RNA polymerase.
FT NON_TER 1
SQ SEQUENCE 977 AA: 107938 MW: 8259A7FBA0191D CRC64;

Query Match
Best Local Similarity 33.3%; Score 120.5; DB 10; Length 977;
Matches 34; Conservative 12; Mismatches 51; Indels 5; Gaps 4;

OY 3 SKGQYPTQPTQYPPG-NPYVPTQTLHLPOAPPTDAPPAVSELYRSPVHGCAATVPTM 61
DB 756 SPGSGSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 11
OY 62 SAAFPGASLYLPMAQSVAV-GPLGSTRIPMAVYVPGPIYPPGS 102
DB 813 PSYSPTSPAVSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPS 854

RESULT 6
O99368 PRELIMINARY: PRT: 1840 AA.
AC 099368;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE DNA-DIRECTED RNA POLYMERASE (EC 2.7.7.6) II LARGEST CHAIN
DE (EC 2.7.7.6).
GN F4B1.70 OR ATAG35800.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Rose M., Hempel S., Entian K.-D., Hohenseil J., Jesse T.,
RA Heijnen L., Vos P., Mewes H.W., Mayer K.F.X., Schueller C.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.

```

```

RA EU Arabidopsis sequencing project;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Rose M., Hempel S., Entian K.-D., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP PARTIAL SEQUENCE FROM N.A.
RA Volckaert G., Grymonprez B., Voet M., Robben J., Mewes H.W.,
RA Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL031986; CAA21466.2;
DR EMBL: AL61588; CAB81489.1;
DR INTERPRO: IPR000684;
DR INTERPRO: IPR000722;
DR INTERPRO: IPR002879;
DR INTERPRO: IPR002965;
DR PFAM: PF00623; RNA_POL_A: 1;
DR PFAM: PF01854; RNA_POL_A2: 1;
DR PRINTS: PRO1217; PRICHEXTENS.
DR PROSITE: PS00115; RNA_POL_II_REPEAT; UNKNOWN_23.
KW DNA-directed RNA polymerase.
SQ SEQUENCE 1840 AA: 204688 MW: 8453621AD945C1B6 CRC64;

Query Match
Best Local Similarity 18.1%; Score 118; DB 10; Length 1840;
Matches 36; Conservative 11; Mismatches 53; Indels 8; Gaps 5;

OY 2 NSKGQYPTQPTQYPPG---PGNPYPTQTLHLPOAPPTDAPPAVSELYRSPVHGCA 55
DB 1583 SSPGSGSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 11
OY 56 ATVPPTMSAFAFGASLYLPMAQSVAV-VGPLGSTRIPMAVYVPGPIYPPGS 102
DB 1642 AVSPPTSPAVSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPS 1689

RESULT 7
O9X208 PRELIMINARY: PRT: 1811 AA.
AC 09X208;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DE 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE LARGEST SUBUNIT OF THE RNA POLYMERASE II COMPLEX (FRAGMENT).
GN RPII215.
OS Drosophila pseudoobscura (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7237;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-99242581; PubMed-10224259;
RA Llopert A., Aguade M.;
RT "Synonymous rates at the RpiI215 gene of Drosophila: variation among
RT species and across the coding region."
RL Genetics 152:269-280(1999).
DR EMBL: Y18879; CAB42797.1;
DR INTERPRO: IPR000684;
DR INTERPRO: IPR000722;
DR INTERPRO: IPR002879;
DR INTERPRO: IPR002965;
DR PFAM: PF00623; RNA_POL_A: 1;
DR PFAM: PF01854; RNA_POL_A2: 1;
DR PRINTS: PRO1217; PRICHEXTENS.
DR PROSITE: PS00115; RNA_POL_II_REPEAT; UNKNOWN_11.
FT NON_TER 1

```


Query Match 18.0%: Score 117.5; DB 5; Length 1889;
 Best Local Similarity 30.5%: Pred. No. 0.027;
 Matches 40; Conservative 10; Mismatches 48; Indels 33; Gaps 7;

QY 8 PTOPTVYVOP---GNVYVQTLHLPOAPPYTDAPPAY---SELYRPSF-----VHP 53
 DB 1758 PTPSPYSPSPSYDSDGSPGPO--YTPGSPQYSPASPKSPSPPLXSPSSPOHSPSNQYSP 1815
 QY 54 GAAIV-----PTMSAAPGASLYLPMASVA-----VGPLGS-TIPMAVYPVGP 96
 DB 1816 TCGTYSATSPRYSPNMSIYSPSTKYSPSPYTPPTARNYSPSPMYSPAPSHYSPSP 1875
 QY 97 YTPGSTVLVE 107
 DB 1876 AYSPSSPTFEE 1886

RESULT 10
 Q9XZS1 PRELIMINARY: PRT; 1889 AA.
 ID 09XZS1
 AC 09XZS1
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
 DE LARGEST SUBUNIT OF THE RNA POLYMERASE II COMPLEX.
 GN RPII215.
 OS Drosophila guanche (Fruit fly).
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxID=7241;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-99242581; PubMed-10224259;
 RA Llopert A.; Aguade M.;
 RT "Synonymous rates at the RPII215 gene of Drosophila: variation among
 RT species and across the coding region."
 RL EMBL: Y18876; CAB42800.1;
 DR EMBL: AF272653; AAF89203.1;
 DR EMBL: AF272643; AAF89193.1;
 DR EMBL: AF272645; AAF89195.1;
 DR EMBL: AF272646; AAF89196.1;
 DR EMBL: AF272647; AAF89197.1;
 DR EMBL: AF272648; AAF89198.1;
 DR EMBL: AF272649; AAF89199.1;
 DR EMBL: AF272650; AAF89200.1;
 DR EMBL: AF272651; AAF89201.1;
 DR EMBL: AF272652; AAF89202.1;
 DR INTERPRO: IPR000684;
 DR INTERPRO: IPR000722;
 DR INTERPRO: IPR002879;
 DR INTERPRO: IPR002965;
 DR PEAM: PF00623; RNA_POL_A; 1.
 DR PFAM: PF01854; RNA_POL_A2; 1.
 DR PRINTS: PR01217; PRICHTEXTENS.
 DR PROSITE: PS00115; RNA_POL_II_REPEAT; UNKNOWN_11.
 SQ SEQUENCE 1889 AA; 209248 MW; D21CBBE9BB586DD0 CRC64;

Query Match 18.0%: Score 117.5; DB 5; Length 1889;
 Best Local Similarity 30.5%: Pred. No. 0.027;
 Matches 40; Conservative 10; Mismatches 48; Indels 33; Gaps 7;

QY 8 PTOPTVYVOP---GNVYVQTLHLPOAPPYTDAPPAY---SELYRPSF-----VHP 53
 DB 1758 PTPSPYSPSPSYDSDGSPGPO--YTPGSPQYSPASPKSPSPPLXSPSSPOHSPSNQYSP 1815
 QY 54 GAAIV-----PTMSAAPGASLYLPMASVA-----VGPLGS-TIPMAVYPVGP 96
 DB 1816 TCGTYSATSPRYSPNMSIYSPSTKYSPSPYTPPTARNYSPSPMYSPAPSHYSPSP 1875
 QY 97 YTPGSTVLVE 107
 DB 1876 AYSPSSPTFEE 1886

RESULT 11
 ID 09XZU9 PRELIMINARY: PRT; 1889 AA.
 AC 09XZU9

DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE LARGEST SUBUNIT OF THE RNA POLYMERASE II COMPLEX (RNA POLYMERASE II
 DE LARGEST SUBUNIT).
 GN RPII215.
 OS Drosophila subobscura (Fruit fly).
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxID=7241;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN-RAICES 111;
 RX MEDLINE-99242581; PubMed-10224259;
 RA Llopert A.; Aguade M.;
 RT "Synonymous rates at the RPII215 gene of Drosophila: variation among
 RT species and across the coding region."
 RL Genetics 152:269-280(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN-VARIOUS STRAINS;
 RX MEDLINE-20341326; PubMed-10880485;
 RA Llopert A.; Aguade M.;
 RT "Nucleotide polymorphism at the RPII215 gene in drosophila subobscura.
 RT Weak selection on synonymous mutations."
 RL Genetics 155:1245-1252(2000).
 DR EMBL: Y18876; CAB42800.1;
 DR EMBL: AF272653; AAF89203.1;
 DR EMBL: AF272643; AAF89193.1;
 DR EMBL: AF272645; AAF89195.1;
 DR EMBL: AF272646; AAF89196.1;
 DR EMBL: AF272647; AAF89197.1;
 DR EMBL: AF272648; AAF89198.1;
 DR EMBL: AF272649; AAF89199.1;
 DR EMBL: AF272650; AAF89200.1;
 DR EMBL: AF272651; AAF89201.1;
 DR EMBL: AF272652; AAF89202.1;
 DR INTERPRO: IPR000684;
 DR INTERPRO: IPR000722;
 DR INTERPRO: IPR002879;
 DR INTERPRO: IPR002965;
 DR PEAM: PF00623; RNA_POL_A; 1.
 DR PFAM: PF01854; RNA_POL_A2; 1.
 DR PRINTS: PR01217; PRICHTEXTENS.
 DR PROSITE: PS00115; RNA_POL_II_REPEAT; UNKNOWN_11.
 SQ SEQUENCE 1889 AA; 209204 MW; 55D013D56DF82A6B CRC64;

Query Match 18.0%: Score 117.5; DB 5; Length 1889;
 Best Local Similarity 30.5%: Pred. No. 0.027;
 Matches 40; Conservative 10; Mismatches 48; Indels 33; Gaps 7;

QY 8 PTOPTVYVOP---GNVYVQTLHLPOAPPYTDAPPAY---SELYRPSF-----VHP 53
 DB 1758 PTPSPYSPSPSYDSDGSPGPO--YTPGSPQYSPASPKSPSPPLXSPSSPOHSPSNQYSP 1815
 QY 54 GAAIV-----PTMSAAPGASLYLPMASVA-----VGPLGS-TIPMAVYPVGP 96
 DB 1816 TCGTYSATSPRYSPNMSIYSPSTKYSPSPYTPPTARNYSPSPMYSPAPSHYSPSP 1875
 QY 97 YTPGSTVLVE 107
 DB 1876 AYSPSSPTFEE 1886

RESULT 12
 ID 096452 PRELIMINARY: PRT; 1553 AA.
 AC 096452
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)

Query Match	17.8%;	Score 116.5;	DB 5;	Length 1553;
Best Local Similarity	31.7%;	Pred No. 0.027;		
Matches 32; Conservative	9;	Mismatches 47;	Indels 13;	Caps 4

RESULT	13	
077165		
ID	077165	PRELIMINARY; PRT; 1690 AA

RC 07-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE DNA-DEPENDENT RNA POLYMERASE II LARGEST SUBUNIT (FRAGMENT).
GN RPB1.
OS Mastigamoeba invertens.
OC Eukaryota; Mastigamoeba.
OX NCBI_TaxID=81100;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC50338;
RA MEDLINE=98426228; Pubmed=9751740;
RX Stillier J.W., Duffield E.C.S., Hall B.D.;
RT "Amtochondriate amoebae and the evolution of DNA-dependent RNA
RT polymerase II.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:11769-11774(1998).
DR EMBL; AF083338; AACG2246.1; -.
DR INTERPRO; IPR000722; -.
DR INTERPRO; IPR002879; -.
DR PFAM; PF00623; RNA_POL_A; 1.
DR PFAM; PF01854; RNA_POL_A2; 1.
FT NON_TER 1 1
SQ SEQUENCE 1690 AA; 184993 MW; 012CDF8B3DD20F64 CRC64;

Query Match	17.7%;	Score 115.5;	DB 5;	Length 1690;
Best Local Similarity	31.1%;	Pred. No. 0.037;		
Matches 32;	Conservative	9;	Mismatches 51;	Indels 11;
			Gaps	4
QY	5	GAYPTQPTV----	PVQPGNPNVYRQTLHLQAPRYTDAPRAYSSELYRPSFVHPGATVPT	60

```

RESULT 14
094231
ID 094231 PRELIMINARY; PRT; 414 AA.
AC 094231;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, last annotation update)
DE RNA POLYMERASE II LARGEST SUBUNIT (FRAGMENT).
OS Kluveromyces lactis.
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales
OC Saccharomycetaceae; Kluveromyces.
OX NCBI_TaxId=28985;
RN [1]
RP SEQUENCE FROM N.A.
RA Lee J.M., Greenleaf A.L.;
RL Submitted (APR-1995) to the EMBL/Genbank/DBJ databases.
DR EMBL: U24217; AAC99803.1; -.
DR INTERPRO: IPR000684; -.
DR INTERPRO: IPR002879; -.
DR INTERPRO: IPR002965; -.
DR PFAM: PF01854; RNA_POL_A2; 1.
DR PRINTS: P01217; PRICHXENSN.
DR PROSITE: PS00115; RNA_POL_II_REPEAT; UNKNOWN_20.
FT NON_TER 1
SQ SEQUENCE 414 AA; 43369 MW; 59A11389369B00B3 CRC64;

```

[illegible]

```

RESULT 15
096446
ID 096446 PRELIMINARY; PRT; 1605 AA.
AC 096446;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE LARGEST SUBUNIT OF RNA POLYMERASE II.
GN RPB1.
OS Valdimorpha necatrix.
OC Eukaryota; Microsporidia; Burenellidae; Valdimorpha.
OX NCBI_TaxID=6039;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99110933; PubMed=9892676;
RA Hirt R.P., Logsdon J.M. JR., Healy B., Dorey M.W., Doolittle W.F.,
  Embley T.M.;
RT "Microsporidia are related to Fungi: evidence from the largest subunit
  of RNA polymerase II and other proteins."
RL Proc. Natl. Acad. Sci. U.S.A. 96:580-585(1999).
DR EMBL; AF060234; AADI2604.1; -.
DR INTERPRO; IPR000684; -.
DR INTERPRO; IPR000722; -.
DR INTERPRO; IPR002879; -.
DR INTERPRO; IPR002965; -.

```

Tue Mar 6 12:02:33 2001

us-09-407-430-1_copy_1_120.rspt

Page 7

DR PFAM: PF00623; RNA_pol_A.1.
DR PFAM: PF01854; RNA_pol_A2.1.
DR PRINTS: PR01217; PRICEXENSN.
DR PROSITE: PS00115; RNA_POL_II_REPEAT; UNKNOWN_17.
SO SEQUENCE 1605 AA; 180964 MW; EDCC0410F528BA2C CRC64;

Query Match	17.3%	Score 113;	DB 5;	Length 1605;
Best Local Similarity	31.4%;	Pred. No. 0.058;		
Matches 32;	Conservative 12;	Mismatches 50;	Indels 8;	Gaps 4;

[illegible]

Search completed: March 6, 2001, 11:54:53
Job time: 250 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 11:51:34 ; Search time 47.68 Seconds
(without alignments)
260.326 Million cell updates/sec

Title: US-09-407-430-2
Perfect score: 2058
Sequence: 1 ETHVYGSAGHTSGFVSL.....ADARVCSCLMMMLISQAEA 363

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues
Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: A_Geneseq_36.*
2: /SIDSI/gcgdata/geneseq/geneseq/AA1980.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq/AA1981.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq/AA1982.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq/AA1983.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq/AA1984.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq/AA1985.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq/AA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2058	100.0	663	17	HCVI E2 + NS2 poly
2	2058	100.0	663	20	Hepatitis C virus
3	2058	100.0	2772	11	Hepatitis C virus
4	2058	100.0	2955	10	Amino acid sequenc
5	2058	100.0	3011	13	Compiled HCV sequ
6	2058	100.0	3011	17	Hepatitis C virus
7	2058	100.0	3011	18	HCV polyprotein.
8	2058	100.0	3011	19	HCV polyprotein.
9	2056	99.9	480	14	HCV-1 E2/NS1 prote
10	2050	99.6	2816	14	HCV-1 polyprotein.
11	2050	99.6	2955	11	Hepatitis C virus
12	2043	99.3	3011	14	Hepatitis C virus

13	2037	99.0	738	14	R33592
14	2036	98.9	402	14	R34438
15	1982	96.3	2894	13	R24440
16	1982	96.3	2894	16	R70230
17	1956	95.0	621	14	R33185
18	1956	95.0	622	14	R33591
19	1955	95.0	409	14	R33995
20	1955	95.0	3011	19	W77398
21	1952	94.8	967	16	W77398
22	1952	94.8	1648	16	R79222
23	1952	94.8	3011	14	R40119
24	1952	94.8	3011	16	R79232
25	1949	94.7	3011	19	W77397
26	1947	94.6	1006	18	W12715
27	1946	94.6	3011	20	W98020
28	1945	94.5	389	14	R40117
29	1945	94.5	537	14	R40114
30	1943	94.4	3011	15	R66995
31	1935	94.0	402	14	R34439
32	1923	93.4	409	14	R33996
33	1911	92.9	3011	13	R22154
34	1903	92.5	402	14	R34440
35	1879	91.3	3011	14	R40120
36	1863	90.5	733	14	R38278
37	1819	88.4	480	14	R33993
38	1817	88.3	353	14	R38089
39	1817	88.3	403	13	R30063
40	1817	88.3	403	20	Y22021
41	1817	88.3	403	20	W75482
42	1815	88.2	367	14	R33994
43	1790	87.0	3011	16	R67588
44	1771	86.1	353	14	R38088
45	1740	84.5	3010	20	W98022

ALIGNMENTS

RESULT	ID	Protein	Length	Score	Alignment
1	R92935	standard: Protein; 663 AA.	663	2058	HCV CKS-full lengt
2	R92935	04-JUN-1996 (first entry)	663	2058	Sequence of glycop
3	R92935	HCVI E2 + NS2 polyptide.	663	2058	Composite HCV HC-3
4	R92935	HCVI E1 protein; E2 protein; truncation; vaccine; diagnosis;	663	2058	Composite heptel
5	R92935	therapy; protein secretion.	663	2058	HCV CKS-NS1-NS1S
6	R92935	Hepatitis C virus.	663	2058	H77 E2/NS1 prote
7	R92935	Hepatitis C virus.	663	2058	Hepatitis C virus-
8	R92935	Hepatitis C virus.	663	2058	PHCVI41-encoded se
9	R92935	Hepatitis C virus.	663	2058	PHCVI76-encoded se
10	R92935	Hepatitis C virus.	663	2058	HCV genomic amino
11	R92935	Hepatitis C virus.	663	2058	HCV sequence. Hep
12	R92935	Hepatitis C virus.	663	2058	Hepatitis C virus
13	R92935	Hepatitis C virus.	663	2058	HCV genome type 1a
14	R92935	Hepatitis C virus.	663	2058	Infectious heptel
15	R92935	Hepatitis C virus.	663	2058	HGH-HCV-E2 fusion
16	R92935	Hepatitis C virus.	663	2058	APP-HCV-E2 fusion
17	R92935	Hepatitis C virus.	663	2058	Hepatitis C virus
18	R92935	Hepatitis C virus.	663	2058	Sequence of glycop
19	R92935	Hepatitis C virus.	663	2058	H90 E2/NS1 prote
20	R92935	Hepatitis C virus.	663	2058	NANV Hutch c59 1s
21	R92935	Hepatitis C virus.	663	2058	Sequence of glycop
22	R92935	Hepatitis C virus.	663	2058	HCV genomic amino
23	R92935	Hepatitis C virus.	663	2058	NANB hepatitis vir
24	R92935	Hepatitis C virus.	663	2058	HCVI E2/NS1 prote
25	R92935	Hepatitis C virus.	663	2058	HCVI E2/NS1 prote
26	R92935	Hepatitis C virus.	663	2058	HCVI E2/NS1 prote
27	R92935	Hepatitis C virus.	663	2058	HCVI E2/NS1 prote
28	R92935	Hepatitis C virus.	663	2058	HCVI E2/NS1 prote
29	R92935	Hepatitis C virus.	663	2058	HCVI E2/NS1 prote
30	R92935	Hepatitis C virus.	663	2058	HCVI E2/NS1 prote
31	R92935	Hepatitis C virus.	663	2058	HCVI E2/NS1 prote
32	R92935	Hepatitis C virus.	663	2058	HCVI E2/NS1 prote
33	R92935	Hepatitis C virus.	663	2058	HCVI E2/NS1 prote
34	R92935	Hepatitis C virus.	663	2058	HCVI E2/NS1 prote
35	R92935	Hepatitis C virus.	663	2058	HCVI E2/NS1 prote
36	R92935	Hepatitis C virus.	663	2058	HCVI E2/NS1 prote
37	R92935	Hepatitis C virus.	663	2058	HCVI E2/NS1 prote
38	R92935	Hepatitis C virus.	663	2058	HCVI E2/NS1 prote
39	R92935	Hepatitis C virus.	663	2058	HCVI E2/NS1 prote
40	R92935	Hepatitis C virus.	663	2058	HCVI E2/NS1 prote
41	R92935	Hepatitis C virus.	663	2058	HCVI E2/NS1 prote
42	R92935	Hepatitis C virus.	663	2058	HCVI E2/NS1 prote
43	R92935	Hepatitis C virus.	663	2058	HCVI E2/NS1 prote
44	R92935	Hepatitis C virus.	663	2058	HCVI E2/NS1 prote
45	R92935	Hepatitis C virus.	663	2058	HCVI E2/NS1 prote

PR 29-JUL-1994; 94US-0282959.
XX (CHIR) CHIRON CORP.
XX Houghton M, Selby M;
XX WPI: 1996-129331/13.
DR N-PSDB: T18246.
XX Secretable hepatitis C virus E1 and E2 polypeptide(s) lacking
PT all/part of the membrane spanning domain - useful in vaccines, and
PT for diagnostic and therapeutic purposes, e.g. in assays for HCV
XX
PS Disclosure; Fig 4a-4c; 46pp; English.
XX Truncated versions of the hepatitis C virus E2 envelope protein
CC (R92935) are obd. by deletion the membrane-spanning C-terminal
CC anchor domain (approx. amino acids 337-361) of the protein.
CC This can be achieved by PCR amplification of the E2 gene (T18246)
CC using appropriate primers. Truncated versions of the E1
CC polypeptide (see R92934) are similarly produced. When produced
CC recombinantly in host cells, the truncated proteins are secreted
CC into the medium. When co-expressed or combined after separate
CC expression, the truncated E2 and E1 proteins form a complex. The
CC truncated proteins and complex are useful in vaccines and for
CC diagnostic and therapeutic applns.
XX
SQ Sequence 663 AA:
Query Match 100.0%; Score 2058; DB 17; Length 663;
Best Local Similarity 100.0%; Pred. No. 4.2e-174;
Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ETHWTGSGAGHTVSGFVSLAPGAKQNVQLINTGSMHLNSTALNCNDSLNTGWLAFY 60
DB 21 ethvrgsgaghtvsgfvsllapgakqnvqlintgsmhlNSTALNCNDSLNTGWLAFY 80
QY 61 HHKFNSSGCPRELAACRPLTDFDQGWGPIISYANGSGPDRPCWHPKPGIYPAKSV 120
DB 81 hnkfnssgcpRELAACRPLTDFDQGWGPIISYANGSGPDRPCWHPKPGIYPAKSV 140
QY 121 GPVYCFPTSPVVGTTDSGAPTYSWGENDVDVFLNTRPPLGNMFCSTWNSGTFTKV 180
DB 141 gpyvcftspvvgtttdsgaptySWGENDVDVFLNTRPPLGNMFCSTWNSGTFTKV 200
QY 181 CGAPPCVIGAGANNLHCPPTDCFRKHDPATYSRCGSGFWITPRCLVDYPRYLMHPTIN 240
DB 201 cgappcvigagannlhcptdcfrkhdpatysrcgsgfwitprclvdypylmhyptcin 260
QY 241 YTIETIRMYVGVEHRLAEACNMWTEGECDEDRRSELSPLLTQTOMOVLPSCFTTLP 300
DB 261 ytifetirmyvgvehrleaacnmwtegercededrseelspllltqtqvlpesfttlp 320
QY 301 ALSTGLIHQNIYDVOYLGVGSSIASMAIKWEYVLLFLLDADARYCSCLMMMLLSIQ 360
DB 321 alstglinhqnivdvqyllyvgvssiasaalkweyvllflldadarycsclmmmlisq 380
QY 361 AEA 363
DB 381 aea 383
RESULT 2
W67615
ID W67615 standard; Protein; 663 AA.
XX W67615;
AC
XX
DT 02-MAR-1999 (first entry)
XX
DE Hepatitis C virus E2 protein.
XX

KW Hepatitis C virus; E2; HCV; truncation; variant; PCR; amplification;
KW affinity chromatography; Galanthus nivalis; agglutinin; medicament;
KW detection; infection.
XX
XX Hepatitis C virus.
OS
XX
XX W09850556-A2.
PN
XX
XX 12-NOV-1998.
PD
XX
XX 06-MAY-1998; 98WO-US09097.
PF
XX
XX 06-MAY-1997; 97US-0045675.
PR
XX
XX (CHIR) CHIRON CORP.
PA
XX
XX Abtignani S, Chien D, Choo QL, Glazer E, Houghton M,
PI Selby M;
XX
XX WPI: 1999-034724/03.
DR N-PSDB: V81371.
DR
XX
XX Methods for isolating truncated HCV E1 and E2 polypeptides - used
PT in, e.g. immunodiagnostic kits for diagnosis of HCV infection
PT
XX
PS Disclosure; Fig 2A-E; 65pp; English.
XX
XX This sequence represents the Hepatitis C virus (HCV) E2 protein. The
CC invention relates to the intracellular production and isolation of
CC C-terminally truncated variants of the E2 protein. The truncations
CC start from about amino acid residue 500 of the E2 protein and are
CC generated by PCR amplification of the gene sequence. The truncated
CC proteins are then produced intracellularly in host cells which are
CC disrupted. The HCV proteins are then purified preferably by affinity
CC chromatography, especially using Galanthus nivalis agglutinin resin.
CC The HCV E1 and HCV E2 polypeptides can be used to manufacture a
CC medicament useful for detecting the presence or absence of HCV infection
CC in an individual. They can also be used in a immunodiagnostic test kit
CC for detecting HCV infection.
XX
SQ Sequence 663 AA:
Query Match 100.0%; Score 2058; DB 20; Length 663;
Best Local Similarity 100.0%; Pred. No. 4.2e-174;
Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ETHWTGSGAGHTVSGFVSLAPGAKQNVQLINTGSMHLNSTALNCNDSLNTGWLAFY 60
DB 21 ethvrgsgaghtvsgfvsllapgakqnvqlintgsmhlNSTALNCNDSLNTGWLAFY 80
QY 61 HHKFNSSGCPRELAACRPLTDFDQGWGPIISYANGSGPDRPCWHPKPGIYPAKSV 120
DB 81 hnkfnssgcpRELAACRPLTDFDQGWGPIISYANGSGPDRPCWHPKPGIYPAKSV 140
QY 121 GPVYCFPTSPVVGTTDSGAPTYSWGENDVDVFLNTRPPLGNMFCSTWNSGTFTKV 180
DB 141 gpyvcftspvvgtttdsgaptySWGENDVDVFLNTRPPLGNMFCSTWNSGTFTKV 200
QY 181 CGAPPCVIGAGANNLHCPPTDCFRKHDPATYSRCGSGFWITPRCLVDYPRYLMHPTIN 240
DB 201 cgappcvigagannlhcptdcfrkhdpatysrcgsgfwitprclvdypylmhyptcin 260
QY 241 YTIETIRMYVGVEHRLAEACNMWTEGECDEDRRSELSPLLTQTOMOVLPSCFTTLP 300
DB 261 ytifetirmyvgvehrleaacnmwtegercededrseelspllltqtqvlpesfttlp 320
QY 301 ALSTGLIHQNIYDVOYLGVGSSIASMAIKWEYVLLFLLDADARYCSCLMMMLLSIQ 360
DB 321 alstglinhqnivdvqyllyvgvssiasaalkweyvllflldadarycsclmmmlisq 380
QY 361 AEA 363
DB 381 aea 383

DB	381	aea	383
RESULT	3		
ID	R08123	standard;	protein; 2772 AA.
XX	AC	R08123;	
XX	DT	23-JAN-1991	(first entry)
XX	DE	Hepatitis C virus polypeptide from long ORF.	
XX	KW	Hepatitis C virus; antiviral agent.	
OS	XX	Hepatitis C virus.	
XX	XX	EP388232-A.	
XX	PD	19-SEP-1990.	
XX	PF	16-MAR-1990;	90EP-0302866.
XX	PR	18-MAY-1989;	89US-0355002.
XX	PR	17-MAR-1989;	89US-0325338.
XX	PR	20-APR-1989;	89US-0341334.
XX	PA	(CHIR-) CHIRON CORP.	
PI	XX	Houghton M, Choo QL, Kuo G;	
DR	XX	WPI: 1990-284418/38.	
DR	XX	N-PSDB: Q05955.	
XX	PT	Hepatitis C virus DNA - used for producing probes,	
XX	PT	polypeptide(s), antibodies and anti-sense polynucleotide(s) for	
XX	PT	diagnosis and therapy.	
PS	XX	Disclousure: Flg 16; 83pp; English.	
XX	CC	HCV cDNA libraries were constructed using pooled serum from a	
CC	CC	chimpanzee with chronic HCV infection. A lambda gtl1 library was	
CC	CC	screened with probes derived from previously isolated clones. The	
CC	CC	ORF is derived from the overlapping clones p14c, CA167b, CA156c,	
CC	CC	CA44c, CA59a, K9-1, 12f, 14i, 11b, 7f, 8h, 33c, 40b, 37b, 35, 36,	
CC	CC	81, 32, 33b, 25c, 14c, 8f, 33f, 33g, 39c, 35f, 19g, 26g and 15e.	
CC	CC	This polypeptide can be used to design probes	
CC	CC	for the detection of HCV nucleic acids., in screening programmes	
CC	CC	for antiviral agents and in preparing blood free of HCV. Antisense	
CC	CC	polynucleotides can be used to inhibit viral replication.	
CC	CC	See also Q05956.	
XX	Sequence	2772 AA;	

	Query Match	100.0%	Score 2058	DB.11	Length 2772
	Best Local Similarity	100.0%	Prod. No. 2.6e-173		
	Matches 363	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	ETHVTSAGCHTVSGFNSLLAPGAKONVOLINTNCSWHLNSTALNCNDSLNTGWLAGEFY	60		
Db	270	ethtvsgaghtltvsgftvaltpgpaqnvqlntngswhlntslatncndslnsgwlaigify	329		
QY	61	HHKFNSSGCPRLRLSCRLTDFDQGWGPISTVANGSGPQORPYCWMHYPPKPGCIYPAKSYVC	120		
Db	330	hhkfnssgcpertlascrpltdfdgsgwprlsgyangsgpqdqrpycwhypkpgciypaksvsc	389		
QY	121	GPVYCFMPSPVVYGTDTDSGAPYISMGENDNDVPLNLTNRPICGWMNCGCTWNSNGCFPKY	180		
Db	390	gpyvcfcpsspvvyltdtsgapylsgwgcndcdvflmnltpdlgwmfgctwmnsgcfkvc	449		
QY	181	CGAPPCVYGGAGNNLTACPTCCFRKHNDATYSRCSGFWITPCLVDYPYRLMHYPCSTIN	240		

Db 450 cgappvcvlgsggnmlhpcidcfcfkhhpaekysrcsgspwltprclvdgpyrlnwhypocln 569

QY 241 YTFIRIKRYVCGVEHRLLEACNMTRGECDELDRNSELSPLLTPTTQWQOVLPDPSPTTTP 3000

Db 510 yllfllrmygvvghrlaeacnwrgercedledrsealsplllttctqvlpcsfctlp 569

QY 301 ALSTGLIHLHONINIVDOVYLKVGSSISMAIKMEYVVLFLFLLLADARVCGCLMMHLLISO 360

Db 570 aistqglhlhqnldvdqylvgvsgstasawlkewyvvllfliladavycscllmmjllisq 6299

QY 361 AEA 363
|||

Db 630 aea 632

Sequence	2955 AA:
XX	Result 4
XX	Y14975
XX	Y14975 standard; Protein: 2955 AA.
XX	Y14975;
XX	08-NOV-1999 (first entry)
XX	Amino acid sequence of HCV-1 ORF.
XX	Hepatitis C virus; HCV; J1; J7; HCV-1; non-A, non-B HCV; NANBH;
XX	HCV infection; vaccine.
XX	Hepatitis C virus.
XX	Key
XX	Location/Qualifiers
XX	Misc-difference 441
XX	FT /note="encoded by T"
XX	Misc-difference 461
XX	FT /note="encoded by CCCC"
XX	EP939128-A2.
XX	01-SEP-1999.
XX	17-SEP-1990; 90EP-0310149.
XX	21-DEC-1989; 89US-0456142.
XX	15-SEP-1989; 89US-0408045.
XX	(CHIR) CHIRON CORP.
XX	(OYAA/) OYA A.
XX	Cha T, Han J, Houghton M, Irvine BD, Kolberg JA;
XX	Miyamura T, Salto I, Weiner AJ;
XX	WPI: 1999-480843/41.
XX	N-PSDB; 207656.
XX	New Hepatitis C Virus isolates, useful for diagnosis of hepatitis
XX	infections and development of vaccines
XX	Disclosure; Fig 12; 132pp; English.
XX	The invention provides two new isolates of hepatitis C virus (HCV), J1
XX	and J7. These two isolates comprise nucleotide and amino acid sequences
XX	that are distinct from the HCV isolate HCV-1. The nucleotide sequences
XX	may be used to detect non-A, non-B HCV (NANBH) polynucleotides by
XX	hybridisation for diagnosis of NANBH infections. They may also be used to
XX	screen blood donors, donated blood and blood products for this infection.
XX	The isolates may also be used to isolate other naturally occurring
XX	variants of the virus. The polypeptides may be used as a vaccine for
XX	administration to patients to protect against infection with NANBH. The
XX	present sequence represents the amino acid sequence of HCV-1 ORF.

Query Match	Similarity	100.0%	Score	2058:	DB	20:	Length	2955:
Best Local	Similarity	100.0%	Pred.	No.	2	8e-173:		
Matches	363:	Conservative	0:	Mismatches	0:	Indels	0:	Gaps
QY	1	ETHWTGSAGHTVSGFVSLAPGAKONVOLINTNGSMHLNSTALNCNDLSLNTGWLAGLFY	60					
Db	384	ethvtygsaghtvsgfvsllapgakvnyllntngswlnstalncondslntgwlaglty	443					
QY	61	HHKFFSSCCPEPLASCRFLTDGCGWCPISYANGSGPDQRPYCMHYPPKPGIIVAKSVY	120					
Db	444	hnhkffssccpeplascrfltdtdggwqplisyangsgpdqrpymhyppkpgiivpaksyvc	503					
QY	121	GPVYCFPTSPVYVGTITDSSGAPTYSGENDTVFLLNNTRPGLGNWFCTMNSTGFTKY	180					
Db	504	gpyvycftspvyvgtitdsgaptysgendtdvfvlnntlrplgmwfgctmnsrgftky	563					
QY	181	CGAPPCVIGGAGNNTLHCPDTCFRKHPRDAPYRRCGSSGPWITRRCIVDPYRLMHYPCITN	240					
Db	564	cgappcvlggagannlthpctdcfrkhpdatysrgssgpltrpcivdpyrlmhyppctln	623					
QY	241	YTFIFRMVYGVGEHRLBACNMWTKGERCDLDDRDRELSPLLLTTQWVLPCEFTTLP	300					
Db	624	ytfifrmvygvgehrlbaacnmwtkgercdledrdreelsplllttqgwvpcefttlp	683					
QY	301	ALSTGLHLHONTVDVQYLYIGVSSIASWAIKWEYVLLFLLADARVCSCLMMMLISQ	360					
Db	684	alstglhlhqnivdvqylyigvssiaswailkweyvvllfllladarycsclmmmlisq	743					
QY	361	AEA 363						
Db	744	aea 746						
RESULT	5							
ID	R21519							
AC	R21519							
XX	22-JUN-1992	(first entry)						
XX	Compiled	HCV sequence.						
XX	HCV1;	serum; gtl1.						
XX	Hepatitis C virus 1.							
XX	Key	Location/Qualifiers						
XX	Misc-difference	9						
XX	Misc-difference	/label= ARG						
XX	Misc-difference	11						
XX	Misc-difference	/label= THR						
XX	Misc-difference	176						
XX	Misc-difference	/label= THR						
XX	Misc-difference	334						
XX	Misc-difference	/label= VAL						
XX	Misc-difference	603						
XX	Misc-difference	/label= ILE						
XX	Misc-difference	848						
XX	Misc-difference	/label= (ASN)						
XX	Misc-difference	1114						
XX	Misc-difference	/label= SER						
XX	Misc-difference	1117						
XX	Misc-difference	/label= THR						
XX	Misc-difference	1276						
XX	Misc-difference	/label= LEU						
XX	Misc-difference	1328						
XX	Misc-difference	/label= (VAL)						
XX	Misc-difference	1454						
XX	Misc-difference	/label= TYR						
XX	Misc-difference	1471						
XX	Misc-difference	/label= (SER)						
XX	Misc-difference	1877						

FT	/label= (GLY)
FT	Misc-difference 1948
FT	/label= (HIS)
FT	Misc-difference 1949
FT	/label= (CYS)
FT	Misc-difference 2021
FT	/label= (VAL)
FT	Misc-difference 2349
FT	/label= (SER)
FT	Misc-difference 2385
FT	/label= (PHE)
FT	Misc-difference 2386
FT	/label= (ALA)
FT	Misc-difference 2502
FT	/label= (PHE)
FT	Misc-difference 2690
FT	/label= (GLY)
FT	Misc-difference 2996
FT	/label= (PRO)
XX	
PX	WO9202642-A.
PD	
XX	20-FEB-1992.
XX	
PF	12-AUG-1991; 91WO-US05728.
PX	
PR	10-AUG-1990; 90US-0566209.
PA	(CHIR-) CHIRON CORP.
PI	Houghton M., Choo QL, Kuo G, Weiner AJ, Urdrea MS, Irvine BD,
PI	Kolberg JA.
DR	WPI, 1992-080094/10.
DR	N-PSDB; Q21744.
XX	
PT	Reagents for isolating, amplifying and detecting HCV
PT	polynucleotide(s) - used to monitor spread of blood-borne non-a,
PT	non-B hepatitis virus infection and screen blood samples for
PT	virus
PS	
XX	Disclosure; Fig 1; 67pp; English.
CC	
CC	Heterogeneities in cloned DNAs of HCV1 are indicated by the amino
CC	acid indicated in the features, the parentheses indicated that the
CC	heterogeneity was detected at or near to the 5'- or 3'-end of the
CC	HCV in the clone.
CC	The sequence is derived from a composite HCV cDNA from HCV1, a
CC	prototypic HCV. The DNA sequence is based upon sequence information
CC	derived from a no. of HCV cDNA clones, which were isolated form a no.
CC	of HCV cDNA libraries, including the "c" library present in lambda
CC	gt11 (ATCC No.40394), and from human serum. The HCV cDNA clones
CC	were isolated by methods described in WO9014436.
CC	The clones from which the sequence was derived are 5'clone32,
CC	b11a, 18g, aq30e, CA205a, CA290a, CA216a, p14a, CA167b, CA156e,
CC	CA84a, CA59a, K9-1 (also called K9-1), 26j, 131, 12f, 141, 11b, 7f,
CC	7e, 8h, 33c, 40b, 37b, 35, 36, 81, 32, 33b, 25c, 14c, 8f, 33f, 33g,
CC	39c, 35f, 19g, 26g, 15e, b5a, 16jh, 6k and pl31jh.
XX	
SQ	Sequence 3011 AA:
	Query Match 100.0%; Score 2058; DB 13; Length 3011;
	Best Local Similarity 100.0%; Pred. No. 2.9e-173;
	Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps
Oy	1 ETHVTTGSGAGHTVSGFVSLLIAPGAKNOLLINTNGSWHLNSTALNCNDSLNTGWLAGLTY 60
Db	384 ethvtgsaghtvsgfvslilapgaknyqlintngswhlntalncnDSLntgwlaGLty 443
Oy	61 HHKFNSSGGPERLARSLRPLTDPDQGCPISYANGSGPPDOPPCOMWHPPPRCGIVPAKSYC 120
Db	444 hnhfnssggperlarSLRpltdpqgwpisYangsgpddqrpyCWHPpprcgIvpaKsyC 503

QY 121 GPVYCTPSPVVVGTTRSGAPTYSWGENDTDFVFLNTRPLGNWFGCTWMNSTGFTKV 180
 Db 504 GPVYCTPSPVVVGTTRSGAPTYSWGENDTDFVFLNTRPLGNWFGCTWMNSTGFTKV 563
 QY 181 CGAPPCVIGGAGNNTLHCPDCKRKHDPDATYSRGSGPWITPRCLVDYPRRLWHYPCTIN 240
 Db 564 CGAPPCVIGGAGNNTLHCPDCKRKHDPDATYSRGSGPWITPRCLVDYPRRLWHYPCTIN 623
 QY 241 YTIKIRMYVGVGVEHRLAEACNWRGRCDEDRDRSELSPLLLTTTQWQVLPCSFITLP 300
 Db 624 YTIKIRMYVGVGVEHRLAEACNWRGRCDEDRDRSELSPLLLTTTQWQVLPCSFITLP 683
 QY 301 ALSTGLIHLHQNIVDVOYLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLWMLLIISQ 360
 Db 684 ALSTGLIHLHQNIVDVOYLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLWMLLIISQ 743
 QY 361 AEA 363
 Db 744 aea 746
 RESULT 6
 R09031
 ID R09031 standard; Protein: 3011 AA.
 XX AC R09031;
 XX DT 15-MAY-1996 (first entry)
 XX Hepatitis C virus polyprotein.
 DE Non-A non-B hepatitis virus; NANBHV; HCV; antigen; detection;
 KW diagnosis; antibodies.
 KW Hepatitis C virus.
 OS
 XX Key Location/Qualifiers
 FT Misc-difference 1..122
 FT /label= antigen
 FT /note= "C22; R09036"
 FT Misc-difference 199..328
 FT /label= antigen
 FT /note= "S2; R09035"
 FT Misc-difference 1192..1457
 FT /label= antigen
 FT /note= "C33c; R09032"
 FT Misc-difference 1569..1931
 FT /label= antigen
 FT /note= "C100; R09033"
 FT Misc-difference 2054..2464
 FT /label= antigen
 FT /note= "NS5; R09034"
 XX EP693687-A1.
 XX 24-JAN-1996.
 XX 03-APR-1991; 91EP-0114016.
 XX 04-APR-1990; 90US-0504352.
 XX (CHIR) CHIRON CORP.
 XX Choo Q, Houghton M, Kuo G;
 XX WPI; 1996-117956/13.
 XX N-PSDB; T12710.
 XX Combinations of synthetic Hepatitis C Virus antigens - provide more
 XX effective diagnosis of Non-A, Non-B Hepatitis
 XX Disclosure; Fig 1(A-Y); 53pp; English.

XX The combination comprises an HCV antigen from the C domain (pref.
 CC C22 - R09036) and at least one HCV antigen from the NS3 (pref. C33c
 CC - R09032), NS4 (pref. C100 - R09033), S (pref. S2 - R09035) or NS5
 CC (R09034) domain.
 CC The antigens may in the form of a fusion protein, a simple physical
 CC mixture, or the individual antigens commonly bound to a solid matrix.
 CC They are pref. prepd. by recombinant DNA techniques (primers are
 CC given in T12711-T12716), but can be synthesised or isolated from
 CC HCV using affinity chromatography.
 XX SQ Sequence 3011 AA;
 Query Match 100.0%; Score 2058; DB 17; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 2.9e-173;
 Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ETHVTGSGAGHTVSGFVSLAPGAKQNVQLINTNGSHLNSTALNCNDSLNTGMLAGLFY 60
 Db 384 ethvtgsgaghtvsgfvsllapgakgnvqlintngshlnstalnclsintgwlglfy 443
 QY 61 HKFNSSGCPERLASCRPLTDFDQGWCPISYANGSGDPDQRPYCHWHYPKPGIVPAKSVC 120
 Db 444 hkhfnssgcpelrascrpltdfdqgwpisyangsgdpdqrpychwhypkpgivpaksvc 503
 QY 121 GPVYCTPSPVVVGTTRSGAPTYSWGENDTDFVFLNTRPLGNWFGCTWMNSTGFTKV 180
 Db 504 GPVYCTPSPVVVGTTRSGAPTYSWGENDTDFVFLNTRPLGNWFGCTWMNSTGFTKV 563
 QY 181 CGAPPCVIGGAGNNTLHCPDCKRKHDPDATYSRGSGPWITPRCLVDYPRRLWHYPCTIN 240
 Db 564 CGAPPCVIGGAGNNTLHCPDCKRKHDPDATYSRGSGPWITPRCLVDYPRRLWHYPCTIN 623
 QY 241 YTIKIRMYVGVGVEHRLAEACNWRGRCDEDRDRSELSPLLLTTTQWQVLPCSFITLP 300
 Db 624 YTIKIRMYVGVGVEHRLAEACNWRGRCDEDRDRSELSPLLLTTTQWQVLPCSFITLP 683
 QY 301 ALSTGLIHLHQNIVDVOYLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLWMLLIISQ 360
 Db 684 ALSTGLIHLHQNIVDVOYLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLWMLLIISQ 743
 QY 361 AEA 363
 Db 744 aea 746
 RESULT 7
 W34480
 ID W34480 standard; Protein: 3011 AA.
 XX AC W34480;
 XX DT 16-MAR-1998 (first entry)
 XX HCV polyprotein.
 DE
 XX PCR primer; amplify; HCV; hepatitis c virus; antigen combination; NS3;
 KW C domain; S domain; NS5; HCV polyprotein; anti-HCV antibody; detection;
 KW NS4.
 XX Hepatitis C virus.
 XX Key Location/Qualifiers
 FH Misc-difference 366
 FT /note= "can optionally be Arg"
 FT Misc-difference 372
 FT /note= "can optionally be Thr"
 FT Misc-difference 867
 FT /note= "can optionally be Thr"
 FT Misc-difference 1341
 FT /note= "can optionally be Val"
 FT Misc-difference 2148

FT Misc-difference /note= "can optionally be Ile" 2883
FT Misc-difference /note= "can optionally be Asn" 3681
FT Misc-difference /note= "can optionally be Ser" 3690
FT Misc-difference /note= "can optionally be Thr" 4167
FT Misc-difference /note= "can optionally be Leu" 4323
FT Misc-difference /note= "can optionally be Val" 4701
FT Misc-difference /note= "can optionally be Tyr" 4752
FT Misc-difference /note= "can optionally be Ser" 5970
FT Misc-difference /note= "can optionally be Gly" 6183
FT Misc-difference /note= "can optionally be His" 6186
FT Misc-difference /note= "can optionally be Cys" 6402
FT Misc-difference /note= "can optionally be Val" 7386
FT Misc-difference /note= "can optionally be Ser" 7494
FT Misc-difference /note= "can optionally be Phe" 7497
FT Misc-difference /note= "can optionally be Ala" 7845
FT Misc-difference /note= "can optionally be Phe" 8409
FT Misc-difference /note= "can optionally be Gly" 9102
FT Misc-difference /note= "can optionally be Gly" 9327
FT Misc-difference /note= "can optionally be Pro"

US5683864-A.
04-NOV-1997.
07-JUL-1992; 92US-0910760.
07-JUL-1992; 92US-0910760.
18-NOV-1987; 87US-0122714.
30-DEC-1987; 87US-0139886.
26-FEB-1988; 88US-0161072.
06-MAY-1988; 88US-0191263.
26-OCT-1988; 88US-0263584.
14-NOV-1988; 88US-0271450.
17-MAR-1989; 89US-0325338.
20-APR-1989; 89US-0341334.
21-APR-1989; 89US-0353896.
04-APR-1990; 90US-0504352.
(CHTR) CHIRON CORP.
Choo Q, Houghton M, Kuo G;
WPI; 1997-548976/50.
N-PSDB; T99981.
Combination of three hepatitis C virus antigens - used for detection
of specific antibodies to diagnose infection
Disclosure; Column 25-46; 57pp; English.

CC This sequence represents the Hepatitis C virus polyprotein. Fragments of
the DNA encoding this sequence can be amplified and used in the
combination of HCV antigens of the invention. The HCV antigen combination
comprises an antigen (Agl) comprising the C domain (i.e. amino acids (aa)
1-120 of the HCV polyprotein), or its immunologically reactive fragment

CC containing at least 8 aa. It also comprises two additional antigens from
two different polyprotein domains, including at least 8 aa from the NS3,
NS4, S or NS5 domains of the polyprotein, corresponding, respectively, to
aa 1050-1640; 1640-2000; 120-400 and 2000-3011 of the HCV polyprotein.
CC Alternatively, Agl contains at least 8 aa from the 1-122 or 9-177 aa
regions of the HCV polyprotein. These antigen combinations are used
diagnostically to detect anti-HCV antibodies, using any standard
immunoassay format. These antigen combinations have a broader range of
reactivity with antibodies than any antigen individually.

XX Sequence 3011 AA;

Query Match 100.0%; Score 2058; DB 18; Length 3011;
Best Local Similarity 100.0%; Pred. No. 2.9e-173;
Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps

QY 1 ETHVTGSGAGTVSGFVSVLLAPGAKONVOLINTNGSWHLNSTALNCNDSINTGWLGLFY 60
DB 384 ethvtgsgagtvsgfvsllapgakvqlntngswhlntalnndsintgwlglfy 443
QY 61 HKFNSGCGPERLASCRPLTDFDQGWGPISYANGSGDPQRPYCWHPKPCGIVPAKSV 120
DB 444 hkfnsngcperlasrpltdfdqgwpiyangsgdpqrpypcwppkpcgivpaksvc 503
QY 121 GPVYCFPTSPVVGTDRSGAPTYSWGENDTDFVLNTRPPPLGNWFGCTWMNSTGFTKV 180
DB 504 gpvycftpspvvgttdrsgaptyswgendtdfvlnttrppplgnwfgctwmnstgftkv 563
QY 181 CGAPPCVTGGAGNNTLHCPDTCFRKHDPATYSRGSGGPWTTPRCLVDYPYRLWHYPC 240
DB 564 cgappcvtggnntlhcpdtdcfrkhdpdatsrgsggptwtprrclvdyprlwhypctin 623
QY 241 YTIKIRMYGVGVEHRLAECNWTGERCDLDRSELSPLLLTTTQWQVLPSCFTTLP 300
DB 624 ytifkirmvgyvgehrleaecnwtgercdledrdselspllltttqvwqlpfcsttllp 683
QY 301 ALSTGLIHLHONIVDVQYLYGVGSSIASWAIKWEYVLLFLLADARVCSCLWMLLSQ 360
DB 684 alstglihlhqnivdvqylyvgvssiaswaikweyvllflladarcvscclwmlllsq 743

QY 361 AEA 363

DB 744 aea 746

RESULT 8
W40038
ID W40038 standard; Protein; 3011 AA.

XX AC W40038;

XX DT 26-MAY-1998 (first entry)

XX DE HCV polyprotein.

XX KW Hepatitis C virus C domain; HCV; C antigen; immunological activity;
NS3 domain; NS4 domain; S domain; NS5 domain.

XX OS Hepatitis C virus.

XX FH Key Location/Qualifiers

FT Domain 1..120

FT Modified-site /label= C_domain

FT /note= "As given in the specification this amino acid can also be Arg"

FT Modified-site 11

FT /note= "As given in the specification this amino acid can also be Thr"

FT Domain 120..400

FT Modified-site /label= S_domain

FT Modified-site 174

FT FT /note- "As given in the specification this amino
 Modified-site 334 acid can also be Thr"
 FT FT /note- "As given in the specification this amino
 Modified-site 603 acid can also be Val"
 FT FT /note- "As given in the specification this amino
 Modified-site 847 acid can also be Ile"
 FT FT /note- "As given in the specification this amino
 Domain 1050..1640 acid can also be Asn"
 FT FT /label= NS3_domain
 Modified-site 1114
 FT FT /note- "As given in the specification this amino
 Modified-site 1217 acid can also be Ser"
 FT FT /note- "As given in the specification this amino
 Modified-site 1276 acid can also be Thr"
 FT FT /note- "As given in the specification this amino
 Modified-site 1328 acid can also be Leu"
 FT FT /note- "As given in the specification this amino
 Modified-site 1452 acid can also be Val"
 FT FT /note- "As given in the specification this amino
 Modified-site 1472 acid can also be Tyr"
 FT FT /note- "As given in the specification this amino
 Domain 1640..2000 acid can also be Ser"
 FT FT /label= NS4_domain
 Modified-site 1877
 FT FT /note- "As given in the specification this amino
 Modified-site 1948 acid can also be Gly"
 FT FT /note- "As given in the specification this amino
 Modified-site 1949 acid can also be His"
 FT FT /note- "As given in the specification this amino
 Domain 2000..3011 acid can also be Cys"
 FT FT /label= NS5_domain
 Modified-site 2021
 FT FT /note- "As given in the specification this amino
 Modified-site 2348 acid can also be Val"
 FT FT /note- "As given in the specification this amino
 Modified-site 2385 acid can also be Ser"
 FT FT /note- "As given in the specification this amino
 Modified-site 2386 acid can also be Phe"
 FT FT /note- "As given in the specification this amino
 Modified-site 2502 acid can also be Ala"
 FT FT /note- "As given in the specification this amino
 Modified-site 2690 acid can also be Phe"
 FT FT /note- "As given in the specification this amino
 Modified-site 2921 acid can also be Gly"
 FT FT /note- "As given in the specification this amino
 Modified-site 2996 acid can also be Gly"
 FT FT /note- "As given in the specification this amino
 acid can also be Pro"

US5712087-A.

27-JAN-1998.

XX

PF 12-MAY-1995; 95US-0440519.
 XX
 PR 07-JUL-1992; 92US-0910760.
 PR 04-APR-1990; 90US-0504352.
 PR 12-MAY-1995; 95US-0440519.
 XX
 PA (CHIR) CHIRON CORP.
 XX
 PI Choo Q, Houghton M, Kuo G;
 XX
 DR WPI; 1998-119973/11.
 DR N-PSDB; V09989.
 XX
 PT Immunoassays for hepatitis C virus antibodies - using combinations
 PT of antigenic fragments of HCV polyprotein
 XX
 PS Disclosure; Fig 1; 59pp; English.
 CC
 CC This sequence represents the hepatitis C virus (HCV) polyprotein which
 CC is used in the construction of novel combinations of HCV antigens that
 CC have a broader range of immunological activity than any single HCV
 CC antigen. An example of such an antigen given in this specification
 CC comprises a first antigen containing at least 8 amino acids of the
 CC C domain of the HCV polyprotein and a second antigen comprising at least
 CC 8 amino acids of the NS3 domain, the NS4 domain, the S domain or the NS
 CC domain of the HCV polyprotein in the form of a fusion protein, a physical
 CC mixture or bound to a solid matrix.
 CC Note: The features given in the specification as represented in the
 CC feature table of W40038 differ from the positions indicated in Figure 1.
 XX
 XX Sequence 3011 AA;
 SQ

Query Match 100.0%; Score 2058; DB 19; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 2.9e-173;
 Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETHVTGSGAGHTVSGFVSLAPGAKONVOLINTNGSWHLNSTALNCNDSLNTGWLGLFY 50
 DB 384 EHTVTGSGAGHTVSGFVSLAPGAKONVOLINTNGSWHLNSTALNCNDSLNTGWLGLFY 50
 QY 61 HHKNSGCGPERLASCRPLTDFDQGWPIYSYANGSGDORPCYCHWYPPKCGIVPAKSV 120
 DB 444 HHKNSGCGPERLASCRPLTDFDQGWPIYSYANGSGDORPCYCHWYPPKCGIVPAKSV 120
 QY 121 GPVYCTPSPVVVGTTRDSGAPTSYWGENDTDFVNLNTRPPLGNWFGTWMNSTGFTKV 180
 DB 504 GPVYCTPSPVVVGTTRDSGAPTSYWGENDTDFVNLNTRPPLGNWFGTWMNSTGFTKV 180
 QY 181 CGAPCVIGGAGNNTLHCPDFCRKHPDATYSRGSGPMTPRCLVDYPRYLWHYPCTIN 240
 DB 564 CGAPCVIGGAGNNTLHCPDFCRKHPDATYSRGSGPMTPRCLVDYPRYLWHYPCTIN 240
 QY 241 YTFIKRMVVGVEHRLAECNWTGRCGLDRLDRSELSPILLTTTQWVLPCSFSTTLP 300
 DB 624 YTFIKRMVVGVEHRLAECNWTGRCGLDRLDRSELSPILLTTTQWVLPCSFSTTLP 300
 QY 301 ALSTGLIHQHNIQVQVLYGVGSSIASMAIKWEYVVLFLLLADARVCSCLMMMLLSQ 360
 DB 684 ALSTGLIHQHNIQVQVLYGVGSSIASMAIKWEYVVLFLLLADARVCSCLMMMLLSQ 360
 QY 361 AEA 363
 DB 744 AEA 746

RESULT 9
 R33992
 ID R33992 standard; Protein: 480 AA.
 XX
 AC R33992;
 XX
 DT 26-JUL-1993 (first entry)

XX HCV-1 E2/NS1 protein.
DE Polymerase chain reaction; PCR; amplify; primer; hepatitis C virus;
XX HCV; asymptomatic; chronically infected; epitope; viral isolate;
KW domain; immunological; cross-reactive; envelope protein; vaccine;
KW gp53(BVDV)/gp55; hog cholera virus; pestivirus; NS1; flavivirus.
XX Synthetic.
OS
XX WO9306126-A.
PN
XX
XX PD
XX 01-APR-1993.
XX 11-SEP-1992; 92WO-US07683.
XX 13-SEP-1991; 91US-0759575.
XX (CHIR) CHIRON CORP.
PA
XX Houghton M, Weiner AJ;
PI
XX WPI; 1993-117468/14.
DR
XX
XX Immuno-reactive hepatitis C virus polypeptide compsns. - contg.
PT at least 2 sequences from the first variable domain of distinct
PT HCV isolates
XX
XX
PS Disclosure; Fig 3; 106pp; English.
XX
XX The sequences given in R33992-002 represent a portion of the E2/NS1
CC protein encoded by group I and group II HCV isolates, from amino acid
CC 370-850. E2/NS1 are viral envelope proteins and are of immunogenic
CC interest. E2/NS1 contains an N-terminal hypervariable domain of about
CC 30 amino acids which shows large variation between nearly all isolates.
CC This is an important immunoreactive domain. This putative envelope
CC glycoprotein E2/NS1 may correspond to the gp53(BVDV)/gp55 (hog cholera
CC virus) envelope polypeptide of the pestiviruses and the NS1 of the
CC flaviviruses, both of which confer protective immunity in hosts
CC vaccinated with these polypeptides. It has been discovered that a
CC number of important HCV epitopes vary among viral isolates and that
CC these epitopes can be mapped to specific domains. This meant that
CC immunologically cross-reactive polypeptides which focus on variable
CC rather than constant domains can be produced. See also Q39134-48
CC and R33982-91.
XX
XX Sequence 480 AA;
SQ
Query Match 99.9%; Score 2056; DB 14; Length 480;
Best Local Similarity 99.7%; Pred. NO. 4.1e-174;
Matches 362; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 ETHVTGSGAGHTVSGFVSLAPGAKONVOLINTNGSWHLNSTALNCNDSLNTGWLGLFY 60
DB 15 ethvtgsgaghtvsgfsvllapgakvnlntngswhlntstalnclnclntgwlglf 74
QY 61 HHKFNSSGCPERLASCRPLTDFDQGWGPIYSANGSGPDORPYCWHYPPKCGIYPAKSV 120
DB 75 hkhfnssgcpelarscrpltdfdqgwgpisyangsgpdprycwhypkpcgipaksv 134
QY 121 GPVYCTFSPVVGTTDRSGAPTSYGENDTDVFLNTRPPLGNWFGCTWNNSTGFTKV 180
DB 135 gpvycftspvvvgtttdrsgaptsygendtdvflntrpplgnwfgctwmnstgftkv 194
QY 181 CGAPPCVIGAGNNTLHCPTDCFRKHPDATYRCGSGGPWITPRCLVDYPRYLWHYPCTIN 240
DB 195 cgappcvigagntlhcptdcfrkhpdatyrcsgsggpwitrclvdypylwhypctin 254
QY 241 YTIKIRMYVGVGVEHLEAACNWTGERCDLEDNRSELSPLLLTTTOMQVLPSCFTTLP 300
DB 255 ytikirmyvgvgehrleaacnwtgercdlednrselelpllltttqvwlpesfttlp 314

QY 301 ALSTGLIHLHONIVDVQYLYGVGSSIASWAIKWEYVVLLELLADARVCSCLMWMLISQ 360
DB 315 alsstglilhqnivdvqylygvgsiaswalkweyvvlfilladarcvscclmmmlisq 374
QY 361 AEA 363
DB 375 aea 377
RESULT 10
R34009
ID R34009 standard; Protein; 2816 AA.
XX
XX AC R34009;
XX 26-JUL-1993 (first entry)
XX HCV-1 polypeptide.
DE
XX
XX Polymerase chain reaction; PCR; amplify; primer; hepatitis C virus;
KW HCV; asymptomatic; chronically infected; epitope; viral isolate;
KW domain; immunological; cross-reactive; envelope protein; vaccine;
KW gp53(BVDV)/gp55; hog cholera virus; pestivirus; NS1; flavivirus.
XX
XX Hepatitis C virus.
OS
XX WO9306126-A.
PN
XX 01-APR-1993.
XX 11-SEP-1992; 92WO-US07683.
XX 13-SEP-1991; 91US-0759575.
XX (CHIR) CHIRON CORP.
XX Houghton M, Weiner AJ;
XX WPI; 1993-117468/14.
XX
XX Immuno-reactive hepatitis C virus polypeptide compsns. - contg.
PT at least 2 sequences from the first variable domain of distinct
PT HCV isolates
XX
XX Disclosure; Fig 9; 106pp; English.
XX
XX This sequence represents the entire hepatitis C virus polypeptide.
CC HCV is a member of the flavivirus family and appears to encode a basic
CC polypeptide domain ("C") at the N-terminal of the viral polypeptide,
CC followed by two glycoprotein domains ("E1", "E2/NS1"), upstream of the
CC nonstructural genes NS2 through NS5. See also Q39134-48, R33982-
CC 4008 and R38088-89.
XX
XX Sequence 2816 AA;
SQ
Query Match 99.6%; Score 2050; DB 14; Length 2816
Best Local Similarity 99.7%; Pred. NO. 1.4e-172;
Matches 362; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ETHVTGSGAGHTVSGFVSLAPGAKONVOLINTNGSWHLNSTALNCNDSLNTGWLGLFY 60
DB 384 ethvtgsgaghtvsgfsvllapgakvnlntngswhlntstalnclnclntgwlglf 443
QY 61 HHKFNSSGCPERLASCRPLTDFDQGWGPIYSANGSGPDORPYCWHYPPKCGIYPAKSV 120
DB 444 hkhfnssgcpelarscrpltdfdqgwgpisyangsgpdprycwhypkpcgipaksv 503
QY 121 GPVYCTFSPVVGTTDRSGAPTSYGENDTDVFLNTRPPLGNWFGCTWNNSTGFTKV 180
DB 504 gpvycftspvvvgtttdrsgaptsygendtdvflntrpplgnwfgctwmnstgftkv 563
QY 181 CGAPPCVIGAGNNTLHCPTDCFRKHPDATYRCGSGGPWITPRCLVDYPRYLWHYPCTIN 240


```

|||||
Db 564 cgappcvlggagntlhcptdcfrkhpdatysrcsgpwiprcldvpyrlwhypctin 623
Qy 241 YTIKIRMYVGGVHRLEAACNWTGRCDEDRSELSPELLTTTQWVLPSCFTTLP 300
Db 624 ytficirmvvgvghrleaacnwtgrecdledrselelplltttqgwvlpcsfittlp 683
Qy 301 ALSTGLIHLHQNIVDVQYLYGVGSSIASWAIKWYVVLFLLLADARVCSCLWMLLIQ 360
Db 684 alstglhlhqnivdvqlyvgvssiaswalkweyvvllfilladarcvscilwmmllisq 743
Qy 361 AEA 363
Db 744 aea 746

RESULT 11
R08124
ID R08124 standard: protein; 2955 AA.
AC R08124;
XX
XX
DT 23-JAN-1991 (first entry)
XX
DE Hepatitis C virus putative polyprotein.
XX
KW Hepatitis C virus (HCV); antiviral agent.
XX
OS Hepatitis C virus.
XX
FH Key Location/Qualifiers
FT Misc-difference 9..9
FT /label=K or R
FT Misc-difference 11..11
FT /label=N or T
FT Misc-difference 176..176
FT /label=I or T
FT Misc-difference 334..334
FT /label=M or V
FT Misc-difference 603..603
FT /label=I or L
FT Misc-difference 848..848
FT /label=Y or N
FT Misc-difference 1114..1114
FT /label=P or S
FT Misc-difference 1117..1117
FT /label-S or T
FT Misc-difference 1276..1276
FT /label=P or L
FT Misc-difference 1454..1454
FT /label=C or Y
FT Misc-difference 1471..1471
FT /label-T or S
FT Misc-difference 1877..1877
FT /label-E or G
FT Misc-difference 1948..1948
FT /label=L or H
FT Misc-difference 1949..1949
FT /label-S or C
FT Misc-difference 2021..2021
FT /label=V or G
FT Misc-difference 2349..2349
FT /label-T or S
FT Misc-difference 2385..2385
FT /label-Y or F
FT Misc-difference 2386..2386
FT /label-S or A
FT Misc-difference 2502..2502
FT /label=L or F
FT Misc-difference 2690..2690
FT /label-R or G
FT Misc-difference 2921..2921
FT /label-R or G

```

```

XX EP388232-A.
PN
XX
XX 19-SEP-1990.
XX
XX 16-MAR-1990; 90EP-0302866.
XX
XX 18-MAY-1989; 89US-0355002.
PR 17-MAR-1989; 89US-0325338.
PR 20-APR-1989; 89US-0341334.
XX
XX (CHIR-) CHIRON CORP.
XX
XX Houghton M, Choo QL, Kuo G;
PI
XX
XX WPI; 1990-284418/38.
DR N-PSDB; Q05956.
XX
XX Hepatitis C virus DNA - used for producing probes,
PT polypeptide(s), antibodies and anti-sense polynucleotide(s) for
PT diagnosis and therapy.
XX
XX Disclosure; Fig 17; 83pp; English.
XX
XX HCV cDNA libraries were constructed using pooled serum from a
CC chimpanzee with chronic HCV infection. A lambda gt10 library was
CC screened with probes derived from previously isolated clones. The
CC ORF is derived from the overlapping clones b114a, ag30a, CA205a,
CC CA290a, CA216a, p14a, CA167b, CA156e, CA84a, CA59a, K9-1, 269, 131,
CC 12f, 14i, 11b, 7f, 8h, 33c, 40b, 37b, 35, 36, 81, 32, 33b, 25c,
CC 14c, 8f, 33f, 33g, 39c, 35f, 19g, 26g, 15e, b5a and 16jh.
CC Polypeptide encoded by this sequence can be used to design probes
CC for the detection of HCV nucleic acids, in screening programmes
CC for antiviral agents and in preparing blood free of HCV. The
CC sequence contains 188 (overlapping) peptides which are claimed as
CC HCV epitopes.
CC See also Q05955.
XX
XX Sequence 2955 AA;
SQ

```

```

Query Match 99.6%; Score 2050; DB 11; Length 2955;
Best Local Similarity 99.4%; Pred. No. 1.4e-172;
Matches 361; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 ETHVTGGSAGHTVSGFVSLLPACAKQNVQLINTNGSWHLNSTALNCNDSLNTGWLGLFY 64
Db 384 ethvtggsaghtvsgfvsllappgkqvqlintngswhlncnclsntgwlglify 443
Qy 61 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGPOORPYCWHYPKPGIVPAKSVC 120
Db 444 hkhfnssgcpelascrpltdfdqgwgpisyangsgpdqrpvcwhypkpgivpakavc 503
Qy 121 GPVYCFPTSPVVVGTTRSGAPTSYGENDTDVFLVNLTRPPLGNHRCCTWNNSTGPTKV 180
Db 504 gpvyctftspvvvgtttrsgaptysygendtdvflntrppdgnhrcctwmnstgptkv 563
Qy 181 CGAPPCVIGGAGNNTLHCPTDCFRKHPDATYSRCGSGPWITPRCLVDYPPYRLWHYPCTIN 240
Db 564 cgappcviggagnntlhcptdcfrkhpdatysrcsgspwxtprclvdypyrwhypctin 623
Qy 241 YTIKIRMYVGGVHRLEAACNWTGRCDEDRSELSPELLTTTQWVLPSCFTTLP 300
Db 624 ytficirmvvgvghrleaacnwtgrecdledrselelplltttqgwvlpcsfittlp 683
Qy 301 ALSTGLIHLHQNIVDVQYLYGVGSSIASWAIKWYVVLFLLLADARVCSCLWMLLIQ 360
Db 684 alstglhlhqnivdvqlyvgvssiaswalkweyvvllfilladarcvscilwmmllisq 743
Qy 361 AEA 363
Db 744 aea 746

```


CC The construction of pHCV-92 (full-length NS1) involved using the
 CC xhoI/BamHI insert from pHCV-78 (see R33590) and ligating that into
 CC the XhoI/BamHI vector backbone of pHCV-80 (see R33591). The resultant
 CC HCV gene represents amino acids 365-847 of the HCV genome. This
 CC resulted in a 1449bp EcoRI/BamHI fragment of HCV cloned into the
 CC CKS fusion vector pJO200. The complete amino acid sequence of the
 CC antigen is designated pHCV-92 (i.e. R33592). The resultant fusion
 CC protein HCV CKS-full length NS1 consists of 239 amino acids of CKS,
 CC seven amino acids contributed by linker DNA sequences and 483 amino
 CC acids from the NS1 region of the HCV genome. The fusion protein is
 CC used to detect antibodies and antigen in body fluids from
 CC individuals exposed to HCV.

XX SQ Sequence 738 AA;

Query Match 99.0%; Score 2037; DB 14; Length 738;
 Best Local Similarity 98.1%; Pred. No. 3.5e-172;
 Matches 362; Conservative 0; Mismatches 1; Indels 6; Gaps 1;

QY 1 ETHVTGSGAGTYSVGFVSLAPGAKONVOLINTNGSWHLNSTALNCNDSINTGWLGLFY 60
 Db 268 ethvtgsgaghtvsgfvslapgakonvqlintngswhlncnclsintgwlglfy 327
 QY 61 HHKFNSSGCGPERLASCRPLTDFDQGWGPISYANGSGDPQRPYCMWHPKPGGIVPAKSVC 120
 Db 328 hkhfnsgcgperlascrpltdfdqggwqisayangsgdpqrpymchwppkpggivpaksvc 387
 QY 121 GPVYCTFTSPVVVGTDRSGAPTSYSGENDTDVFLNTRPPLGNFPGCTWMNSTGFTKV 180
 Db 388 gpvyctfptspvvvgttdrsgaptsyngendtdvflntrpplgnwfgctwmnstgftkv 447
 QY 181 CGA-----PPCVIGGAGNNTLCPTDCFRKHPDATYSRCGSGPWITPRLCLVDYPRYLWH 234
 Db 448 cgappcvigppcvigagntlhcptdcfrkhpdatysrcsgspwltprclvdypylwh 507
 QY 235 YPTINTYTFKIMYVGGVEHRLAECNWTGERCDLEDRLDRSELSPLLLTTTQWQVLPC 294
 Db 508 yptintytfkirmyvvgvvehrleaacnwtgercdledrldrselspllltttqwqlpc 567
 QY 295 SFTTLPALSTGLHLHQNIVDVQYLYGVGSSIASWAIKWEYVLLFLLADARVCSCLWM 354
 Db 568 sfttlpalstgllhlhqnivdvqylyvgvssiaswaikweyvllflladavrcsclwm 627
 QY 355 MLLISQAEA 363
 Db 628 mllisqaea 636

RESULT 14
 R34438
 ID R34438 standard; Protein; 402 AA.
 XX AC R34438;

DT 09-AUG-1993 (first entry)

DE Sequence of glycoprotein E2/NS1 in clone HCV1.
 XX Hepatitis C virus; envelope protein; glycoprotein; E2/NS1;
 KW diagnostic reagent.

OS Hepatitis C virus.

XX EP537626-A.

PN 21-APR-1993.

XX 08-OCT-1992; 92EP-0117191.

XX 08-OCT-1991; 91JP-0260824.

XX (NAHE-) NAT INST OF HEALTH.

XX PI Harada S, Honda Y, Miyamura T, Saito I;
 XX WPI; 1993-127516/16.
 DR N-PSDB; Q40330.
 XX Diagnostic reagent for hepatitis C virus - comprises second
 PT envelope protein or first non-structural protein encoded by HCV
 PT gene and has sugar chain
 XX Claim 2; Pages 30-32; 58pp; English.
 CC Glycoprotein E2/NS1 is derived from the second envelope protein or
 CC first non-structural protein encoded by the genome of HCV. The
 CC nucleic acid is extracted from the serum of the patient of hepatitis
 CC C. The serum is pref. mixed with transfer RNA (tRNA) as a carrier
 CC of virus RNA. As a technique of cloning cDNA from the nucleic acid,
 CC it is preferred to use polymerase chain reaction method. In the
 CC reaction, any commercially available random primers or synthesized
 CC DNA having a base sequence similar to that of primer NS1 may be used
 CC as a primer. Representative examples of sense primers includes S1.
 XX SQ Sequence 402 AA;

Query Match 98.9%; Score 2036; DB 14; Length 402;
 Best Local Similarity 100.0%; Pred. No. 2e-172;
 Matches 358; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETHVTGSGAGTYSVGFVSLAPGAKONVOLINTNGSWHLNSTALNCNDSINTGWLGLFY 60
 Db 45 ethvtgsgaghtvsgfvslapgakonvqlintngswhlncnclsintgwlglfy 104
 QY 61 HHKFNSSGCGPERLASCRPLTDFDQGWGPISYANGSGDPQRPYCMWHPKPGGIVPAKSVC 120
 Db 105 hkhfnsgcgperlascrpltdfdqggwqisayangsgdpqrpymchwppkpggivpaksvc 164
 QY 121 GPVYCTFTSPVVVGTDRSGAPTSYSGENDTDVFLNTRPPLGNFPGCTWMNSTGFTKV 180
 Db 165 gpvyctfptspvvvgttdrsgaptsyngendtdvflntrpplgnwfgctwmnstgftkv 224
 QY 181 CGAPPVIGGAGNNTLCPTDCFRKHPDATYSRCGSGPWITPRLCLVDYPRYLWHYPCCTIN 240
 Db 225 cgappcvigagntlhcptdcfrkhpdatysrcsgspwltprclvdypylwhypcctin 284
 QY 241 YTFIKIRMTYVGGVEHRLAECNWTGERCDLEDRLDRSELSPLLLTTTQWQVLPCSFETTL 300
 Db 285 ytfikirmyvvgvvehrleaacnwtgercdledrldrselspllltttqwqlpcsfettp 344
 QY 301 ALSTGLIHLHQNIVDVQYLYGVGSSIASWAIKWEYVLLFLLADARVCSCLWMMLLI 358
 Db 345 alstgllhlhqnivdvqylyvgvssiaswaikweyvllflladavrcsclwmmlli 402

RESULT 15
 R24440
 ID R24440 standard; protein; 2894 AA.
 XX AC R24440;

DT 02-DEC-1992 (first entry)

DE Composite HCV HC-J1/CDC/CHI protein.

XX Hepatitis C virus; peptides; antibodies; ELISA.

XX Synthetic.

XX Key Location/Qualifiers

FT Peptide 1..20 /label= 1

FT Peptide 7..26 /label= 2

FT Peptide 13..32
FT /label= 3
FT Peptide 37..56
FT /label= 4
FT Peptide 49..68
FT /label= 5
FT Peptide 61..80
FT /label= 6
FT Peptide 73..92
FT /label= 7
FT Peptide 168..1707
FT /label= 8
FT Peptide 1694..1713
FT /label= 9
FT Peptide 1706..1725
FT /label= 10
FT Peptide 1712..1731
FT /label= 11
FT Peptide 1718..1737
FT /label= 12
FT Peptide 1724..1743
FT /label= 13
FT Peptide 1730..1749
FT /label= 14
FT Peptide 2263..2282
FT /label= 15
FT Peptide 2275..2294
FT /label= 16
FT Peptide 2287..2306
FT /label= 17
FT Peptide 2299..2318
FT /label= 18
FT Peptide 2311..2330
FT /label= 19
XX
XX EP489968-A.
XX
XX 17-JUN-1992.
XX
XX 14-DEC-1990; 90EP-0124241.
XX
XX 14-DEC-1990; 90EP-0124241.
XX (INNO-) INNOGENETICS NV.
XX
XX Deleys RJ, Maertens G, Pollet D, Van Heuverswyn H;
XX
XX WPI; 1992-201383/25.
XX
XX New synthetic peptides for detecting antibodies to hepatitis C
XX virus - useful in e.g. ELISA assays, and for detection of HCV
XX antigens or as immunogens
XX
XX Disclosure; Fig 1; 32pp; English.

XX RNA viruses frequently exhibit a high rate of spontaneous mutation,
XX thus a virus is considered to be the same of equiv. to HCV if it
XX exhibits a global homology of more than 70 percent with the HCV HC-
XX J1/CDC/CHI composite sequence. the peptide fragments of this DNA
XX sequence indicated in the features table can immunologically mimic
XX proteins encoded by HCV. Additional amino acids or chemical gps.
XX may be added to either end of the peptides for the purpose of
XX creating a linker arm for attachment to a carrier. The peptides can
XX be used for the detection of antibodies specific for HCV. They may
XX be used in the form of kits, opt. with reagents such as
XX staphylococcal protein A, streptococcal protein G, avidin or
XX streptavidin. The peptides may also be used as immunogens for
XX raising antibodies.

XX Sequence 2894 AA;

Query Match 96.3%; Score 1982; DB 13; Length 2894;

Best Local Similarity 95.9%; Pred. No. 1.5e-166;
Matches 348; Conservative 5; Mismatches 10; Indels 0; Gaps 0;
QY 1 ETHVTGSGAGHTVSGFVSLIAPCAKONVOLINTNGSHLNSTALNCNDSLNTGWLACLFY 60
Db 384 etivsggqaaramsgivslftpgakgnliqintngshlnstalnncnesintgwlaliy 443
QY 61 HHKFNSSGCPERLASCRPLTDFDQGWGCPISYANGSGPDQRPYCHWHYPPKPCGIVPAASVC 120
Db 444 qhkfnsngcperlasrpltdfdqggwgpisyangsgpdqrpchwhypkpcgivpaksvc 503
QY 121 GPVYCFPSVVGTTDRSGAPTYSMGENDTVFVLNTRPPLGNWFGCTWMNSTGFTKV 180
Db 504 gpvycftpsvvgtttdrsgaptyswgendtdvflntrpplgnwfgctwmnstgftkv 563
QY 181 CGAPPCVIGGAGNNTLHCPTDCPRKHPDATYSRGSGSPWITPRCLVDYDYLWHYPCTIN 240
Db 564 cgappcviggagnntlhcptdcprkhpdatysrcsgspwitprclvdypylwhypctin 623
QY 241 YTIKIRMYGGVEHRLAEAAACNMTRGERCDLEDRSELSPLLLITTTQWQVLPSCFTTLP 300
Db 624 ytifkirmvgygvehrleaaacnmtrgercdleedrseisplllittqwgvlpcsfittlp 683
QY 301 ALSTGLIHLHQNIVDVQYLYGVGSSSTASWAKWEYVVLLELLADARVCSLWMLLIQS 360
Db 684 alstgllhlhqnivdvqylyvgvssstaswawkweyvvllfllladarvcslwmlliqs 743
QY 361 AEA 363
Db 744 aea 746

Search completed: March 6, 2001, 11:51:45
Job time: 68 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 11:53:04 ; Search time 45.68 Seconds
(without alignments)
539.578 Million cell updates/sec

Title: US-09-407-430-2

Perfect score: 2058

Sequence: 1 ETHVGTGSAGHTVSGFVSL.....ADRVCSCLWMMLLSQAEA 363

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 6790655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_66:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	2058	100.0	3011	1 GNMVC3	genome polyprotein
2	1943	94.4	3011	1 S40770	genome polyprotein
3	1926	93.6	3011	1 GNMVCH	genome polyprotein
4	1817	88.3	716	2 JQ1366	polyprotein - hepa
5	1717	83.4	3010	1 GNMVTC	genome polyprotein
6	1710	83.1	3010	1 GNMVTC	genome polyprotein
7	1706	82.9	3010	1 GNMVTC	genome polyprotein
8	1696	82.4	782	2 S19875	genome polyprotein
9	1687	82.0	782	2 S19876	genome polyprotein
10	1674	81.3	782	2 S18032	genome polyprotein
11	1673	81.3	3010	1 S18030	genome polyprotein
12	1672	81.2	782	2 S18031	genome polyprotein
13	1671	81.2	3010	1 A45573	genome polyprotein
14	1651.5	80.2	3014	1 JC5620	genome polyprotein
15	1641	79.7	787	2 PN0677	hypothetical prote
16	1555	75.6	3033	1 GNMVJ8	genome polyprotein
17	1531	74.4	3033	1 JQ1303	genome polyprotein
18	1458	70.8	350	2 S35631	genome polyprotein
19	1325	64.4	640	2 JQ1584	genome polyprotein
20	1001	48.6	234	2 S32742	genome polyprotein
21	966	46.9	235	2 S32747	genome polyprotein
22	942	45.8	237	2 S32744	genome polyprotein
23	806	39.2	415	2 PC4407	genome polyprotein
24	689	33.5	513	2 A44150	envelope protein - structural protein
25	688.5	33.5	876	2 PC2219	polypeptide - hepa
26	623	30.3	138	2 S24080	envelope protein -
27	609	29.6	138	2 S24081	envelope protein -
28	572	27.8	138	2 S24074	envelope protein -
29	568	27.6	138	2 S24075	envelope protein -

ALIGNMENTS

RESULT 1

GNMVC3

genome polyprotein - hepatitis C virus (strain HCV-1)

N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstru protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 26-May-2000

C:Accession: A39166; PQ0403; PQ0404

R:Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.;

Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991

A:Title: Genetic organization and diversity of the hepatitis C virus.

A:Reference number: A39166; MUID:91172826

A:Accession: A39166

A:Molecule type: mRNA

A:Residues: 1-3011 <CHO>

A:Cross-references: GB:M62321; NID:g329873; PTDN:AAA45676.1; PTD:g329874

R:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap,

J. Gen. Virol. 73, 1131-1141, 1992

A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship t

A:Reference number: PQ0393; MUID:92268871

A:Accession: PQ0403

A:Molecule type: genomic RNA

A:Residues: 1577-1633 <CHA>

A:Cross-references: DBJ:DI0128

A:Experimental source: isolates E-b16

A:Accession: PQ0404

A>Status: preliminary

A:Molecule type: genomic RNA

A:Residues: 1577-1633 <CH2>

A:Experimental source: isolates E-b17

C:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology

C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstru

F:1-115/Product: capsid protein C #status predicted <CPC>

F:116-191/Product: envelope protein M #status predicted <EPM>

F:192-389/Product: major envelope protein E #status predicted <MEE>

F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>

F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

F:1007-1615/Product: hepacivirin #status predicted <NS3>

F:1230-1493/Domain: DEAD/H box helicase homology <DEAD>

F:1312-1317/Region: nucleotide-binding motif A (P-loop)

F:1316-1319/Region: DXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4A>

F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4B>

F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077

Query Match

Best Local Similarity 100.0%; Score 2058; DB 1; Length 3011;

Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 338; Conservative 12; Mismatches 13; Indels 0; Gaps 0;

QY 1 ETHVTGSGAGHTVSGFVSLAPGAKQNVOLINTNGSWHLNSTALNCNDSLNTGWLGLFY 60
Db 384 ETHVTGSGAGHTVSGFVSLAPGAKQNVOLINTNGSWHLNSTALNCNDSLNTGWLGLFY 443

QY 61 HHKFNSSGCGPERLASCRPLTDFDQGWGPISYANGSGDORPCYCHWYPPKPGCVIPAKSV 120
Db 444 QHKNSSGCGPERLASCRPLTDFDQGWGPISYANGSGDORPCYCHWYPPKPGCVIPAKSV 503

QY 121 GPVYCFTPSPVVVGTDRSGAPTSYSGENDTDVFLNNTRPPLGNWFGCTWNNSTGFTKV 180
Db 504 GPVYCFTPSPVVVGTDRSGAPTSYSGANDTDVFLNNTRPPLGNWFGCTWNNSTGFTKV 563

QY 181 CGAPPCVIGAGNNTLHCPTDCFRKHDPDATYSCGSGPWTTPRCLVDYPRYLWHYPCTIN 240
Db 564 CGAPPCVIGAGNNTLHCPTDCFRKHDPDATYSCGSGPWTTPRCLVDYPRYLWHYPCTIN 623

QY 241 YTIKIRMYVGGVEHRLAEACNWTGRCDCLEDRDRSELSPLLLTTTQWVLPKCSFTTLP 300
Db 624 YTIKIRMYVGGVEHRLAEACNWTGRCDCLEDRDRSELSPLLLTTTQWVLPKCSFTTLP 683

QY 301 ALSTGLIHLHQNIVDVQYLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLMWLLISQ 360
Db 684 ALSTGLIHLHQNIVDVQYLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLMWLLISQ 743

QY 361 AEA 363
Db 744 AEA 746

RESULT 4
JQ1366
polyprotein - hepatitis C virus (French isolate) (fragments)
C:Species: hepatitis C virus
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 29-May-1998
C:Accession: JQ1366
R:Kremersdorf, D.; Porchon, C.; Kim, J.P.; Reyes, G.R.; Brechot, C.
J. Gen. Virol. 72, 2557-2561, 1991
A:Title: Partial nucleotide sequence analysis of a French hepatitis C virus: implication
A:Reference number: JQ1366; MUID:92013977
A:Accession: JQ1366
A:Molecule type: genomic RNA
A:Residues: 1-716 <KRE>
C:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology
C:Keywords: glycoprotein; envelope protein; glycoprotein; DEAD/H box helicase homology
F:84,90,97,115,143,199,223,243,290,312/Binding site: carbohydrate (Asn) (covalent) #stat

Query Match 88.3%; Score 1817; DB 2; Length 716;
Best Local Similarity 86.7%; Pred. No. 9.9e-137;
Matches 306; Conservative 30; Mismatches 17; Indels 0; Gaps 0;

QY 1 ETHVTGSGAGHTVSGFVSLAPGAKQNVOLINTNGSWHLNSTALNCNDSLNTGWLGLFY 60
Db 51 ETVYTCGSGARTTQGLVSLFSGAKQDIQINTNGSWHLNSTALNCNDSLNTGWLGLFY 110

QY 61 HHKFNSSGCGPERLASCRPLTDFDQGWGPISYANGSGDORPCYCHWYPPKPGCVIPAKSV 120
Db 111 YHKNSSGCGPERLASCRPLTDFDQGWGPISYANGSGDORPCYCHWYPPKPGCVIPAKSV 170

QY 121 GPVYCFTPSPVVVGTDRSGAPTSYSGENDTDVFLNNTRPPLGNWFGCTWNNSTGFTKV 180
Db 171 GPVYCFTPSPVVVGTDRSGAPTSYSGENDTDVFLNNTRPPLGNWFGCTWNNSTGFTKV 230

QY 181 CGAPPCVIGAGNNTLHCPTDCFRKHDPDATYSCGSGPWTTPRCLVDYPRYLWHYPCTIN 240
Db 231 CGAPPCVIGAGNNTLHCPTDCFRKHDPDATYSCGSGPWTTPRCLVDYPRYLWHYPCTIN 290

QY 241 YTIKIRMYVGGVEHRLAEACNWTGRCDCLEDRDRSELSPLLLTTTQWVLPKCSFTTLP 300
Db 291 YTIKIRMYVGGVEHRLAEACNWTGRCDCLEDRDRSELSPLLLTTTQWVLPKCSFTTLP 350

QY 301 ALSTGLIHLHQNIVDVQYLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLMW 353
Db 351 ALSTGLIHLHQNIVDVQYLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLMW 403

RESULT 5
GNVTC
genome polyprotein - hepatitis C virus
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstru
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 26-May-2000
C:Accession: A38465
R:Takamiyawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E
J. Virol. 65, 1105-1113, 1991
A:Title: Structure and organization of the hepatitis C virus genome isolated from hum
A:Reference number: A38465; MUID:91140698
A:Accession: A38465
A:Molecule type: genomic RNA
A:Residues: 1-3010 <TA>
A:Cross-references: EMBL:M58335; NID:g329770; PIDN:AAA72945.1; PID:g329771
C:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruct
F:2-115/Product: capsid protein C #status predicted <CP>
F:116-191/Product: envelope protein M #status predicted <EPW>
F:192-389/Product: major envelope protein E #status predicted <ME>
F:390-709/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome polyprotein NS3 #status predicted <NS3>
F:1230-1493/Domain: DEAD/H box helicase homology <DEAD>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077

Query Match 83.4%; Score 1717; DB 1; Length 3010;
Best Local Similarity 80.4%; Pred. No. 4.2e-128;
Matches 292; Conservative 29; Mismatches 42; Indels 0; Gaps 0;

QY 1 ETHVTGSGAGHTVSGFVSLAPGAKQNVOLINTNGSWHLNSTALNCNDSLNTGWLGLFY 60
Db 384 ETHVTGSGAGHTVSGFVSLAPGAKQNVOLINTNGSWHLNSTALNCNDSLNTGWLGLFY 443

QY 61 HHKFNSSGCGPERLASCRPLTDFDQGWGPISYANGSGDORPCYCHWYPPKPGCVIPAKSV 120
Db 444 THSNSSGCGPERMAOQRTTDFDQGWGPISYANGSGDORPCYCHWYPPKPGCVIPASEVC 503

QY 121 GPVYCFTPSPVVVGTDRSGAPTSYSGENDTDVFLNNTRPPLGNWFGCTWNNSTGFTKV 180
Db 504 GPVYCFTPSPVVVGTDRSGAPTSYSGENDTDVFLNNTRPPLGNWFGCTWNNSTGFTKV 563

QY 181 CGAPPCVIGAGNNTLHCPTDCFRKHDPDATYSCGSGPWTTPRCLVDYPRYLWHYPCTIN 240
Db 564 CGPPCNIGGVGNNTLTCPTDCFRKHDPDATYSCGSGPWTTPRCLVDYPRYLWHYPCTIN 623

QY 241 YTIKIRMYVGGVEHRLAEACNWTGRCDCLEDRDRSELSPLLLTTTQWVLPKCSFTTLP 300
Db 624 YTIKIRMYVGGVEHRLAEACNWTGRCDCLEDRDRSELSPLLLTTTQWVLPKCSFTTLP 683

QY 301 ALSTGLIHLHQNIVDVQYLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLMWLLISQ 360
Db 684 ALSTGLIHLHQNIVDVQYLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLMWLLISQ 743

QY 361 AEA 363
Db 744 AEA 746

RESULT 6

GNMWTV

genome polyprotein - hepatitis C virus (strain Taiwan)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain Taiwan)
C:Keywords: ATP; capsid protein C; envelope protein M; hepatitis C virus (strain Taiwan)
F:1-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: major envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein M #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome: sequence determination and mapping the
F:1230-1493/Domain: hepatitis C virus genome: sequence determination and mapping the
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F:196, 209, 233, 234, 250, 305, 325, 417, 423, 430, 448, 532, 556, 576, 623, 645, 1213, 1255, 2041, 207

C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 26-May-2000
C:Accession: A40244
R:Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.
Virology 188, 102-113, 1992
A:Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the
A:Reference number: A40244; MUID:92230206
A:Accession: A40244
A:Molecule type: genomic RNA
A:Residues: 1-3010 <CHE>
A:Cross-references: GB:M84754
C:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology
C:Keywords: ATP; capsid protein C; envelope protein M; hepatitis C virus (strain Taiwan)
F:1-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: major envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein M #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome: sequence determination and mapping the
F:1230-1493/Domain: hepatitis C virus genome: sequence determination and mapping the
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F:196, 209, 233, 234, 250, 305, 325, 417, 423, 430, 448, 532, 556, 576, 623, 645, 1213, 1255, 2041, 207

Query Match 83.1%; Score 1710; DB 1; Length 3010;
Best Local Similarity 79.6%; Pred. No. 1.5e-127;
Matches 288; Conservative 33; Mismatches 41; Indels 0; Caps 0;

Qy 2 THVTGSGAGHTVSGFVSLAPGAKONVOLINTNGSWHLNSTALNCNDSLTGWLGLFYH 61
Db 385 TIVSGTIVARTHSASLFTGASOKIQLINTNGSWHLNSTALNCNDSLTGWLGLFYH 444
Qy 62 HKFNSSGCPERLASCRPLTDFDQGWGPIYANGSGDPQPCWHPKPGCIVPAKSVCG 121
Db 445 HRFNAGSGPERMASCRPIDFQAQGWGPIYANGSGDPQPCWHPKPGCIVPAKSVCG 504
Qy 122 PVYCFTPSPVVGTTDRSGAPTSWGENDTDFVFLNTRPPLGNWFGCTWNNSTGFTKVC 181
Db 505 PVYCFTPSPVVGTTDRSGAPTSWGENETDVLILNTRPPLGNWFGCTWNNSTGFTKVC 564
Qy 182 GAPPVIGGAGNNTLHCPTDCFRKHDPATYSRCGSGPWITPRCLVDYPIRLWHYPCTINY 241
Db 565 GPPCNIGGGNNTLVCPDCEFRKHDPATYSRCGSGPWITPRCMVDYPIRLWHYPCTVNF 624
Qy 242 TIFKIRMYVGGVHRLAECNWTGERCDLEDRDRSELSPLLLTTQWQVLPSCFTTLPA 301
Db 625 TIFKIRMYVGGVHRLAECNWTGERCDLEDRDRSELSPLLLTTQWQVLPSCFTTLPA 684
Qy 302 LSTGLIHLHQNIVDQYLYGVGSSSTASWAIKWEYVLLFLLLADARVCSCLWMLLSQA 361
Db 685 LSTGLIHLHQNIVDQYLYGVGSSSTASWAIKWEYVLLFLLLADARVCSCLWMLLSQA 744
Qy 362 EA 363
Db 745 EA 746

RESULT 7

GNMVCJ

genome polyprotein - hepatitis C virus (strain J)
N:Contains: capsid protein C; envelope protein M; major envelope protein E; nonstructural
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus

C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 16-Jun-2000
C:Accession: A39253; PS0086
R:Kato, N.; Hijikata, M.; Ootsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Sugimura, T.; Shi
Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990
A:Title: Molecular cloning of the human hepatitis C virus genome from Japanese patient
A:Reference number: A39253; MUID:91088550
A:Accession: A39253
A:Molecule type: genomic RNA
A:Residues: 1-3010 <KAT>
A:Cross-references: GB:D90208; NID:g221610; PIDN:BAA14233.1; PID:g221611
R:Kato, N.; Ohkoshi, S.; Shimotohno, K.
Proc. Jpn. Acad. 65B, 219-223, 1989
A:Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence v

A:Reference number: PS0085
A:Accession: PS0086
A:Molecule type: genomic RNA
A:Residues: 2650-2707 <KAT>
A:Experimental source: Japanese isolate
C:Comment: The cleavage sites of this polyprotein have not been determined.
C:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology
C:Keywords: ATP; glycoprotein; hydrolase; P-loop; polyprotein; serine proteinase; tra
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: major envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome: sequence determination and mapping the
F:1230-1493/Domain: hepatitis C virus genome: sequence determination and mapping the
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F:196, 209, 234, 250, 305, 325, 417, 423, 430, 448, 532, 556, 576, 623, 645, 1213, 1255, 2041, 2077, 224

Query Match 82.9%; Score 1706; DB 1; Length 3010;
Best Local Similarity 78.7%; Pred. No. 3.1e-127;
Matches 285; Conservative 39; Mismatches 38; Indels 0; Caps 0;

Qy 2 THVTGSGAGHTVSGFVSLAPGAKONVOLINTNGSWHLNSTALNCNDSLTGWLGLFYH 61
Db 385 THVTGGRVASTQSLVSWLSQPSKIQLVNTNGSWHLNSTALNCNDSLTGWLGLFYH 444
Qy 62 HKFNSSGCPERLASCRPLTDFDQGWGPIYANGSGDPQPCWHPKPGCIVPAKSVCG 121
Db 445 HRFNAGSGPERMASCRPIDFQAQGWGPIYANGSGDPQPCWHPKPGCIVPAKSVCG 504
Qy 122 PVYCFTPSPVVGTTDRSGAPTSWGENDTDFVFLNTRPPLGNWFGCTWNNSTGFTKVC 181
Db 505 PVYCFTPSPVVGTTDRSGAPTSWGENETDVLILNTRPPLGNWFGCTWNNSTGFTKVC 564
Qy 182 GAPPVIGGAGNNTLHCPTDCFRKHDPATYSRCGSGPWITPRCLVDYPIRLWHYPCTINY 241
Db 565 GPPCNIGGGNNTLVCPDCEFRKHDPATYSRCGSGPWITPRCMVDYPIRLWHYPCTVNF 624
Qy 242 TIFKIRMYVGGVHRLAECNWTGERCDLEDRDRSELSPLLLTTQWQVLPSCFTTLPA 301
Db 625 TIFKIRMYVGGVHRLAECNWTGERCDLEDRDRSELSPLLLTTQWQVLPSCFTTLPA 684
Qy 302 LSTGLIHLHQNIVDQYLYGVGSSSTASWAIKWEYVLLFLLLADARVCSCLWMLLSQA 361
Db 685 LSTGLIHLHQNIVDQYLYGVGSSSTASWAIKWEYVLLFLLLADARVCSCLWMLLSQA 744
Qy 362 EA 363
Db 745 EA 746

RESULT 8

S19875

genome polyprotein - hepatitis C virus (isolate JK3) (fragment)

N:Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein
C:Species: hepatitis C virus
A:Variety: isolate JK3
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 26-Aug-1999
C:Accession: S19875
R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
submitted to the EMBL Data Library, September 1991
A:Description: Sequence analysis of putative structural regions of Hepatitis C Virus iso
A:Reference number: S18029
A:Accession: S19875
A:Molecule type: genomic RNA
A:Residues: 1-782 <HON>
A:Cross-references: EMBL:X61592; NID:g59482; PIDN:CAA43789.1; PID:g59483
A:Experimental source: isolate JK3
C:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology
C:Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural
F:1-191/Product: core protein #status predicted <MAT1>
F:192-383/Product: envelope protein 1 #status predicted <MAT2>
F:384-733/Product: NS1/E2 protein #status predicted <MAT3>
F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 82.4%; Score 1696; DB 2; Length 782;
Best Local Similarity 78.2%; Pred. No. 4.5e-127;
Matches 284; Conservative 38; Mismatches 41; Indels 0; Gaps 0;

QY 1 ETHVTGSGAGTYSVGFVSLAPCAKONVOLINTNGSHLNSTALNCNDSLNTGWLGLFY 60
Db 384 QTRVTAQVGRITSSLTSPGPSONIQLVNSGSHINRTALSCNDSLKTGLAALFY 443
QY 61 HHKFNSSGCPERLASRPLTDFDQGWPIYSANGSDQRPYCHWYPPKPGCIVPAKSYVC 120
Db 444 THKFNASGCPERMASCRSIDTDFDQGWPIHVPNTDQKPYCHWYAPRPGCIVPAKSYVC 503
QY 121 GPVYCTPSPVVVGTTRDSCAPYISGENDTDVFLNTRPPLGNMFGCTWMNSTGFTKV 180
Db 504 GPVYCTPSPVVVGTTRDSCAPYISGENDTDVFLNTRPPLGNMFGCTWMNSTGFTKV 563
QY 181 CGAPPCVIGGAGNTLHCPTDCFRKHPDATYSRCSGSPWITPRCLVDYPRVRLWHYPCTIN 240
Db 564 CGGPPCNIGGAGNTLHCPTDCFRKHPDATYSRCSGSPWITPRCLVDYPRVRLWHYPCTIN 623
QY 241 YTIKIRMYVGGVEHRLAECNWTGERCDLDRSELSPLLLTTTQWQVLPSCFTTLP 300
Db 624 FSVKIRMYVGGVEHRLAECNWTGERCDLDRSELSPLLLTTTQWQVLPSCFTTLP 683
QY 301 ALSTGLIHLHONIVDQYLYGVGSSSTASNAIKWEYVVLFLLLADARVCSCLMMMLLSQ 360
Db 684 ALSTGLIHLHONIVDQYLYGVGSSSTASNAIKWEYVVLFLLLADARVCSCLMMMLLSQ 743
QY 361 AEA 363
Db 744 AEA 746

Query Match 82.4%; Score 1696; DB 2; Length 782;
Best Local Similarity 78.2%; Pred. No. 4.5e-127;
Matches 284; Conservative 38; Mismatches 41; Indels 0; Gaps 0;

QY 1 ETHVTGSGAGTYSVGFVSLAPCAKONVOLINTNGSHLNSTALNCNDSLNTGWLGLFY 60
Db 384 QTRVTAQVGRITSSLTSPGPSONIQLVNSGSHINRTALSCNDSLKTGLAALFY 443
QY 61 HHKFNSSGCPERLASRPLTDFDQGWPIYSANGSDQRPYCHWYPPKPGCIVPAKSYVC 120
Db 444 THKFNASGCPERMASCRSIDTDFDQGWPIHVPNTDQKPYCHWYAPRPGCIVPAKSYVC 503
QY 121 GPVYCTPSPVVVGTTRDSCAPYISGENDTDVFLNTRPPLGNMFGCTWMNSTGFTKV 180
Db 504 GPVYCTPSPVVVGTTRDSCAPYISGENDTDVFLNTRPPLGNMFGCTWMNSTGFTKV 563
QY 181 CGAPPCVIGGAGNTLHCPTDCFRKHPDATYSRCSGSPWITPRCLVDYPRVRLWHYPCTIN 240
Db 564 CGGPPCNIGGAGNTLHCPTDCFRKHPDATYSRCSGSPWITPRCLVDYPRVRLWHYPCTIN 623
QY 241 YTIKIRMYVGGVEHRLAECNWTGERCDLDRSELSPLLLTTTQWQVLPSCFTTLP 300
Db 624 FSVKIRMYVGGVEHRLAECNWTGERCDLDRSELSPLLLTTTQWQVLPSCFTTLP 683
QY 301 ALSTGLIHLHONIVDQYLYGVGSSSTASNAIKWEYVVLFLLLADARVCSCLMMMLLSQ 360
Db 684 ALSTGLIHLHONIVDQYLYGVGSSSTASNAIKWEYVVLFLLLADARVCSCLMMMLLSQ 743
QY 361 AEA 363
Db 744 AEA 746

RESULT 9
S19876
genome polyprotein - hepatitis C virus (isolate JK5) (fragment)
N:Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein
C:Species: hepatitis C virus
A:Variety: isolate JK5
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 26-Aug-1999
C:Accession: S19876
R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
submitted to the EMBL Data Library, September 1991
A:Description: Sequence analysis of putative structural regions of Hepatitis C Virus iso
A:Reference number: S18029
A:Accession: S19876
A:Molecule type: genomic RNA
A:Residues: 1-782 <HON>
A:Cross-references: EMBL:X61595; NID:g59486; PIDN:CAA43792.1; PID:g59487
A:Experimental source: isolate JK5
C:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology

C:Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural
F:1-191/Product: core protein #status predicted <MAT1>
F:192-383/Product: envelope protein 1 #status predicted <MAT2>
F:384-733/Product: NS1/E2 protein #status predicted <MAT3>
F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 82.0%; Score 1687; DB 2; Length 782;
Best Local Similarity 78.8%; Pred. No. 2.3e-126;
Matches 286; Conservative 34; Mismatches 43; Indels 0; Gaps 0;

QY 1 ETHVTGSGAGTYSVGFVSLAPCAKONVOLINTNGSHLNSTALNCNDSLNTGWLGLFY 60
Db 384 QTRVTAQVGRITSSLTSPGPSONIQLVNSGSHINRTALSCNDSLKTGLAALFY 443
QY 61 HHKFNSSGCPERLASRPLTDFDQGWPIYSANGSDQRPYCHWYPPKPGCIVPAKSYVC 120
Db 444 THKFNASGCPERMASCRSIDTDFDQGWPIHVPNTDQKPYCHWYAPRPGCIVPAKSYVC 503
QY 121 GPVYCTPSPVVVGTTRDSCAPYISGENDTDVFLNTRPPLGNMFGCTWMNSTGFTKV 180
Db 504 GPVYCTPSPVVVGTTRDSCAPYISGENDTDVFLNTRPPLGNMFGCTWMNSTGFTKV 563
QY 181 CGAPPCVIGGAGNTLHCPTDCFRKHPDATYSRCSGSPWITPRCLVDYPRVRLWHYPCTIN 240
Db 564 CGGPPCNIGGAGNTLHCPTDCFRKHPDATYSRCSGSPWITPRCLVDYPRVRLWHYPCTIN 623
QY 241 YTIKIRMYVGGVEHRLAECNWTGERCDLDRSELSPLLLTTTQWQVLPSCFTTLP 300
Db 624 FTVKIRMYVGGVEHRLAECNWTGERCDLDRSELSPLIRATTEQVILPCSTTLP 683
QY 301 ALSTGLIHLHONIVDQYLYGVGSSSTASNAIKWEYVVLFLLLADARVCSCLMMMLLSQ 360
Db 684 ALSTGLIHLHONIVDQYLYGVGSSSTASNAIKWEYVVLFLLLADARVCSCLMMMLLSQ 743
QY 361 AEA 363
Db 744 AEA 746

RESULT 10
S18032
genome polyprotein - hepatitis C virus (isolate JK4) (fragment)
N:Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein
C:Species: hepatitis C virus
A:Variety: isolate JK4
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 29-May-1998
C:Accession: S18032
R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
submitted to the EMBL Data Library, September 1991
A:Description: Sequence analysis of putative structural regions of Hepatitis C Virus
A:Reference number: S18029
A:Accession: S18032
A:Molecule type: genomic RNA
A:Residues: 1-782 <HON>
A:Cross-references: EMBL:X61594
A:Experimental source: isolate JK4
C:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology
C:Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural
F:1-191/Product: core protein #status predicted <MAT1>
F:192-383/Product: envelope protein 1 #status predicted <MAT2>
F:384-733/Product: NS1/E2 protein #status predicted <MAT3>
F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 81.3%; Score 1674; DB 2; Length 782;
Best Local Similarity 78.7%; Pred. No. 2.5e-125;
Matches 285; Conservative 33; Mismatches 44; Indels 0; Gaps 0;

QY 2 THVTGSGAGTYSVGFVSLAPCAKONVOLINTNGSHLNSTALNCNDSLNTGWLGLFY 61
Db 385 TTVSGHASQITRGVTSFSPGSAQIQLVNTNGSHINRTALNCNDSINTGFFAALFYA 444

QY 182 GAPPVIGGAGNNTLHCPDTCFRKHPDATYRCGSGPWITPRCLVDYPRYLWHYPCITNY 241
 Db 565 EGPCNIGGVCNNLTCTDTCFRKHPDATYTKCGSGPMLTPRCMVHYPRYLWHYPCITVNF 624

QY 242 TIFKIRMYGVGVEHRLAECANWTRGERCDLDRDRSELSPLLLTTTQOVLPSCFTTLPA 301
 Db 625 TIFKIRMYGVGVEHRLAECANWTRGERCDLDRDRSELSPLLLTTTQOVLPSCFTTLPA 684

QY 302 LSTGLIHLHQNIVDVQVLYGVGSSIASWAIKWEYVLLFLLLDARVCSCIMMMLLISOA 361
 Db 685 LSTGLIHLHQNIVDVQVLYGVGSSIASWAIKWEYVLLFLLLDARVCACIMMMLLISOA 744

QY 362 EA 363
 Db 745 EA 746

RESULT 13
 A45573
 genome polyprotein - hepatitis C virus (strain JT)
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain JT)
 C:Species: hepatitis C virus
 C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 16-Jun-2000
 C:Accession: A45573
 R:Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata, Y.
 A:Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier: S
 A:Reference number: A45573; MUID:92295714
 A:Accession: A45573
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-3010 <N>
 A:Cross-references: GB:D11168; GB:D01171; NID:g221612; PID:BAA01943.1; PID:g221613
 A:Experimental source: HCV-JT
 A:Note: sequence extracted from NCBI backbone (NCBI:106206, NCBI:106207)
 C:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology
 F:115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <MEP>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: hepatitis C virus #status predicted <NS3>
 F:1230-1493/Product: DEAD/H box helicase homology <DEAD>
 F:1230-1237/Product: nucleotide-binding motif A (P-loop)
 F:1312-1317/Product: nucleotide-binding motif B
 F:1316-1319/Product: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4>
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 81.2%; Score 1671; DB 1; Length 3010;
 Best Local Similarity 78.5%; Pred. No. 1,9e-124;
 Matches 284; Conservative 29; Mismatches 49; Indels 0; Gaps 0;

QY 2 THVTGSGAGHTVGVFSLAPAGAKQVQLINTNGSHWLNSTALNCNDSLNTGLAGLYFH 61
 Db 385 THVTGSGAGHTVGVFSLAPAGAKQVQLINTNGSHWLNSTALNCNDSLNTGLAGLYFH 444

QY 62 HKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGP-DORPCWHYPPKPGCVIPAKSVCG 121
 Db 445 HKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGP-DORPCWHYPPKPGCVIPAKSVCG 504

QY 122 PVYCFPTSPVVGTTDRSGAPYSWENDTDVFLNTRPPLNPNFCTWNNSTGFTKVC 181
 Db 505 PVYCFPTSPVVGTTDRSGAPYSWENDTDVFLNTRPPLNPNFCTWNNSTGFTKVC 564

QY 182 GAPPVIGGAGNNTLHCPDTCFRKHPDATYRCGSGPWITPRCLVDYPRYLWHYPCITNY 241
 Db 565 EGPCNIGGVCNNLTCTDTCFRKHPDATYTKCGSGPMLTPRCMVHYPRYLWHYPCITVNF 624

QY 242 TIFKIRMYGVGVEHRLAECANWTRGERCDLDRDRSELSPLLLTTTQOVLPSCFTTLPA 301
 Db 625 TIFKIRMYGVGVEHRLAECANWTRGERCDLDRDRSELSPLLLTTTQOVLPSCFTTLPA 684

QY 302 LSTGLIHLHQNIVDVQVLYGVGSSIASWAIKWEYVLLFLLLDARVCSCIMMMLLISOA 361
 Db 685 LSTGLIHLHQNIVDVQVLYGVGSSIASWAIKWEYVLLFLLLDARVCACIMMMLLISOA 744

QY 362 EA 363
 Db 745 EA 746

RESULT 14
 JC5620
 genome polyprotein - hepatitis C virus (isolate EUH1480)
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus (isolate EUH1480)
 C:Species: hepatitis C virus
 C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-May-2000
 C:Accession: JC5620
 R:Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.
 A:Title: The complete coding sequence of hepatitis C virus genotype 5a, the predomina
 A:Reference number: JC5620; MUID:97366593
 A:Accession: JC5620
 A:Molecule type: mRNA
 A:Residues: 1-3014 <CHA>
 A:Cross-references: GB:Y13184
 A:Experimental source: genotype 5a, which predominates in South Africa
 A:Note: the translation of the nucleotide sequence is not complete in this paper
 C:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology
 C:Keywords: ATP; glycoprotein; hydrolase; P-loop; polyprotein; serine proteinase; tra
 F:2-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <MEP>
 F:384-408/Product: hypervariable #status predicted
 F:730-730/Product: nonstructural protein NS1 #status predicted <NS1>
 F:731-1007/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1008-1616/Product: hepatitis C virus #status predicted <NS3>
 F:1231-1494/Product: DEAD/H box helicase homology <DEAD>
 F:1231-1238/Product: nucleotide-binding motif A (P-loop)
 F:1313-1318/Product: nucleotide-binding motif B
 F:1317-1320/Product: DEXH motif
 F:1617-1863/Product: nonstructural protein NS4a #status predicted <NS4>
 F:1864-2014/Product: nonstructural protein NS4b #status predicted <NS4>
 F:2015-3014/Product: nonstructural protein NS5 #status predicted <NS5>
 F:2210-2249/Product: interferon sensitivity determining #status predicted

Query Match 80.2%; Score 1651.5; DB 1; Length 3014;
 Best Local Similarity 75.8%; Pred. No. 6.7e-123;
 Matches 275; Conservative 39; Mismatches 48; Indels 1; Gaps 1;

QY 2 THVTGSGAGHTVGVFSLAPAGAKQVQLINTNGSHWLNSTALNCNDSLNTGLAGLYFH 61
 Db 385 THVTGSGAGHTVGVFSLAPAGAKQVQLINTNGSHWLNSTALNCNDSLNTGLAGLYFH 444

QY 62 HKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGP-DORPCWHYPPKPGCVIPAKSVCG 120
 Db 445 HKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGP-DORPCWHYPPKPGCVIPAKSVCG 504

QY 121 GPVYCFPTSPVVGTTDRSGAPYSWENDTDVFLNTRPPLNPNFCTWNNSTGFTKVC 180
 Db 505 GPVYCFPTSPVVGTTDRSGAPYSWENDTDVFLNTRPPLNPNFCTWNNSTGFTKVC 564

QY 181 GAPPVIGGAGNNTLHCPDTCFRKHPDATYRCGSGPWITPRCLVDYPRYLWHYPCITNY 240
 Db 565 GAPPVIGGAGNNTLHCPDTCFRKHPDATYRCGSGPWITPRCLVDYPRYLWHYPCITNY 624

QY 241 YTIKIRMYGVGVEHRLAECANWTRGERCDLDRDRSELSPLLLTTTQOVLPSCFTTLPA 300
 Db 565 YTIKIRMYGVGVEHRLAECANWTRGERCDLDRDRSELSPLLLTTTQOVLPSCFTTLPA 624

Db 625 YTIKVRMFIGGLEHRLAEACNWTYGERCDLEDRLDRSPLLLHTTQWAILPCSFPTTP 684

Qy 301 ALSTGLIHQHNVQVLYGVGSSIASWAKWEYVVLFLLLADARVCSCLLMMLLISQ 360

Db 685 ALSTGLIHQHNVQVLYGVGSSIASWAKWEYVVLFLLLADARVCSCLLMMLLISQ 744

Qy 361 AEA 363

Db 745 AEA 747

RESULT 15

PN0677

hypothetical protein 787 - hepatitis C virus (fragment)

C:Species: hepatitis C virus

C:Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 26-Aug-1999

C:Accession: PN0677

R:Cho, S.H.; Yoon, J.I.; Chang, J.E.; Ahn, B.M.; Lee, C.H.; Lee, Y.I.

Biochem. Biophys. Res. Commun. 196, 780-788, 1993

A:Title: Genomic typing of hepatitis C viruses from Korean patients: Implications of gen

A:Reference number: PN0677; MUID:94059104

A:Accession: PN0677

A:Molecule type: mRNA

A:Residues: 1-787 <CHO>

A:Cross-references: GB:L20498; NID:g1381031; PIDN:AAB02608.1; PID:g1381032

C:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology

C:Keywords: glycoprotein; nonstructural protein

F:196,209,234,250,305,325,421,427,452,536,544,560,580,627,649/Binding site: carbohydrate

Query Match 79.7%; Score 1641; DB 2; Length 787;

Best Local Similarity 77.4%; Pred. No. 1e-122;

Matches 281; Conservative 32; Mismatches 50; Indels 0; Gaps 0;

Qy 1 EPHVTGSGAGHTVSGFVLLAPGAKONVOLINTNGSWHLNSTALNCNDSLTGWLGLFY 60

Db 388 DTRVTGAQAGRTTQGFSLFMPGQKIQLIHTNGSWHLNSTALNCNDSLTGFLAALFY 447

Qy 61 HHKFNSSGCPERLASCRPLTDGCGPITSYANGSDPDQPCYCHYPPKPGCIVPAKSVC 120

Db 448 THSFNSSGGPERMAOCRPIIDFAQQGWGLTHDASGNLDQRPYCSSLAPKPGCIVPASQVC 507

Qy 121 GPVYCFTSPVVVGTDRSGAPTSWGENDTDVFLNNTRPPLGNWFGCTWMNSTGFTKV 180

Db 508 GPVYCFTSPVVVGTDRFGVPTYSWGANETDVLNNTLPQGNWFGCTWMNGTGSTKT 567

Qy 181 CGAPPCVGGAGNNLHCPTDCFRKHDPATYRCGSGPWITPRCLVDYPRLWHYPCTIN 240

Db 568 CGRPCKIGGVGNNTLICPTDCFRKHPEATYTKCGSGPWLPRCMVDYPRLWHYPCTVN 627

Qy 241 YTIKIRMYGVGVEHRLAEACNWTGERCDLEDRLDRSPLLLHTTQWAILPCSFPTTP 300

Db 628 FSVFKVRMIVGVGEORLNACNWTGERCDLEDRLDRSPLLLHTTQWAILPCSFPTTP 687

Qy 301 ALSTGLIHQHNVQVLYGVGSSIASWAKWEYVVLFLLLADARVCSCLLMMLLISQ 360

Db 688 ALSTGLIHQHNVQVLYGVGSSIASWAKWEYVVLFLLLADARVCACLLMMLLVAQ 747

Qy 361 AEA 363

Db 748 AEA 750

Search completed: March 6, 2001, 11:53:15

Job time: 152 sec

DR PFAM; PF01560; HCV_NS1; 1.
DR PFAM; PF01538; HCV_NS2; 1.
DR PFAM; PF01006; HCV_NS4a; 1.
DR PFAM; PF01001; HCV_NS4b; 1.
DR PFAM; PF01506; HCV_NS5a; 1.
DR PFAM; PF00998; HCV_RDRP; 1.
DR PFAM; PF01543; HCV_capsid; 1.
DR PFAM; PF01542; HCV_core; 1.
DR PFAM; PF01539; HCV_env; 1.
KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
FT INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 729
FT CHAIN 730 1006
FT CHAIN 1007 1615
FT CHAIN 1616 1862
FT CHAIN 1863 2013
FT CHAIN 2014 3011
FT TRANSMEM 347 369
FT ACT_SITE 1083 1083
FT ACT_SITE 1107 1107
FT ACT_SITE 1165 1165
FT NP_BIND 1230 1237
FT SITE 1316 1319
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 234 234
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 476 476
FT CARBOHYD 532 532
FT CARBOHYD 540 540
FT CARBOHYD 556 556
FT CARBOHYD 576 576
FT CARBOHYD 623 623
FT CARBOHYD 645 645
FT CARBOHYD 2041 2041
FT CARBOHYD 2077 2077
FT CARBOHYD 2240 2240
FT CARBOHYD 2364 2364
FT CARBOHYD 2789 2789
SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;

Query Match 100.0%; Score 2058; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 5.8e-167;
Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETHVTCGAGTGVGVSVLLAPGAKONVLTNGSWHLNSTALNCNDSLNTGWLGLFY 60
DB 384 ETHVTCGAGTGVGVSVLLAPGAKONVLTNGSWHLNSTALNCNDSLNTGWLGLFY 443

QY 61 HHKFNSSGGPERLASCRPLTDFDQGWGPISYANGSGPDQRPYCHWYPPKCGIVPAKVC 120
DB 444 HHKFNSSGGPERLASCRPLTDFDQGWGPISYANGSGPDQRPYCHWYPPKCGIVPAKVC 503

QY 121 GPVYCTFPVVGTTDRSGATYSWGENDTDVFLNNTRPPLGNWGCWTWNSGFTKTV 180
DB 504 GPVYCTFPVVGTTDRSGATYSWGENDTDVFLNNTRPPLGNWGCWTWNSGFTKTV 563

QY 181 CGAPPCVIGAGNNTLHCPTDCFRKHPDATYRCGSGPWITPRCLVDYDYPRLWHYPCCTIN 240
DB 564 CGAPPCVIGAGNNTLHCPTDCFRKHPDATYRCGSGPWITPRCLVDYDYPRLWHYPCCTIN 623

QY 241 YTIKIRMYVGGVEHLEACNWTGRCERCDLEDRSELSPLLLTTTQWVLPCTTLP 300

DB 624 YTIKIRMYVGGVEHLEACNWTGRCERCDLEDRSELSPLLLTTTQWVLPCTTLP 683
QY 301 ALSTGLIHQNTVDVQYLYGVGSSIASWAIKWEYVYVLLFLLADARVCSCLMMMLLSQ 360
DB 684 ALSTGLIHQNTVDVQYLYGVGSSIASWAIKWEYVYVLLFLLADARVCSCLMMMLLSQ 743
QY 361 AEA 363
DB 744 AEA 746

RESULT 2
POLG_HCVH STANDARD; PRT; 3011 AA.
ID POLG_HCVH
AC P27958;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN) (P22);
DE ENVELOPE GLYCOPROTEIN E1 (GP32); ENVELOPE GLYCOPROTEIN E2
DE (GP68) (GP70) (NS1); PROTEIN P7; NONSTRUCTURAL PROTEIN NS2 (P21)
DE (EC 3.4.22.-); PROTEASE/HELICASE NS3 (P70) (EC 3.4.21.-);
DE NONSTRUCTURAL PROTEIN NS4A (P4); NONSTRUCTURAL PROTEIN NS4B (P27);
DE NONSTRUCTURAL PROTEIN NS5A (P56); NONSTRUCTURAL PROTEIN NS5B (P66)
DE (P70) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate H) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE-92052256; PubMed-1659800;
RX Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
RA Prince A.M.;
RT "Genomic structure of the human prototype strain H of hepatitis C
RT virus: comparison with American and Japanese isolates.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
[2]
RN X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
RP MEDLINE-97331322; PubMed-9187654;
RX Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
RT "Structure of the hepatitis C virus RNA helicase domain.";
RL Nat. Struct. Biol. 4:463-467(1997).
[3]
RN X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
RP MEDLINE-98154321; PubMed-9493270;
RX Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
RA Murcko M.A., Lin C., Caron P.R.;
RT "Hepatitis C virus NS3 RNA helicase domain with a bound
RT oligonucleotide: the crystal structure provides insights into the mode
RT of unwinding.";
RL Structure 6:89-100(1998).
CC -1- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS3.
CC -1- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
CC -1- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
CC ACTIVATION OF NS3.
CC -1- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
CC -1- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
CC -1- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
CC -1- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY C18.
CC -1- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY C19.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. usage by and for commercial

RT virus: a 2.2-A resolution structure in a hexagonal crystal form.";
 RL Protein Sci 7:837-847(1998).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL: M58335; AAA72945.1; -
 PIR: A38465; GNWVTC.
 PDB: 1AIQ; 25-MAR-98.
 PDB: 1JXP; 14-JAN-98.
 PDB: 1NS3; 08-APR-98.
 MEROPS: S29.001; -
 MEROPS: 039.001; -
 INTERPRO: IPR000745; -
 INTERPRO: IPR001490; -
 INTERPRO: IPR002166; -
 INTERPRO: IPR002518; -
 INTERPRO: IPR002519; -
 INTERPRO: IPR002521; -
 INTERPRO: IPR002522; -
 INTERPRO: IPR002531; -
 INTERPRO: IPR002868; -
 PFAM: PF01560; HCV_NS1; 1.
 PFAM: PF01538; HCV_NS2; 1.
 PFAM: PF01006; HCV_NS4a; 1.
 PFAM: PF01001; HCV_NS4b; 1.
 PFAM: PF01506; HCV_NS5a; 1.
 PFAM: PF00998; HCV_RdrP; 1.
 PFAM: PF01543; HCV_capsid; 1.
 PFAM: PF01542; HCV_core; 1.
 PFAM: PF01539; HCV_env; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430

FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 3010 AA; 327189 MW; F8422D5ECCDFD9C CRC64;

Query Match 83.4%; Score 1717; DB 1; Length 3010;

Best Local Similarity 80.4%; Pred. No. 6e-138;

Matches 292; Conservative 29; Mismatches 42; Indels 0; Gaps 0;

QY 1 ETHVTGSGAGTSGVFYSLAPGAKQNVQLINTNGSHLNTALNCNDSLNTGWLGLFY 60
 DB 384 DTHVTGGAQAKTTRNLVSMFASGSPQKIQLINTNGSHINRTALNCNDSLQTFGLAIFY 443
 QY 61 HHKFNSSGCPERLASRPLTDFDQWGPISYANGSGDPQRPYCHYPPKPGIIVPAKSVC 120
 DB 444 THSFNSSGCPERMAQCRITDKFDQWGPITYAESRSDQRPYCHYPPPOCTIVPASEVC 503
 QY 121 GPVYCFPTSPVVGTTDRSGAPTYSWGNEDTVFLNNTTRPLGNWFCCTWMNSTGFTKV 180
 DB 504 GPVYCFPTSPVVGTTDRSGAPTYSWGNEDTVFLNNTTRPLGNWFCCTWMNSTGFTKV 563
 QY 181 CGAPPCVIGGAGNNTLHCPTDCFRKHPDATYSRCGSGPWITPCLVDYPRHLHYPTIN 240
 DB 564 CGGPPCNIGVGNNLTCTPTDCFRKHPDATYSRCGSGPWITPCLVDYPRHLHYPTIN 623
 QY 241 YTIKIRMYGVGVEHRLAENACNWTNRGERCDLEDRDRSELPLLLTQWQVLPSCFTLP 300
 DB 624 FTIFKVRMYGVGVEHRLAENACNWTNRGERCDLEDRDRSELPLLLTQWQVLPSCFTLP 683
 QY 301 ALSTGLHLHONIVDQVLYGVGSSIASWAIKWEYVLLFLLADARVCSLWMLLIQ 360
 DB 684 ALSTGLHLHONIVDQVLYGVGSSIASWAIKWEYVLLFLLADARVCSLWMLLIQ 743
 QY 361 AEA 363
 DB 744 AEA 746

RESULT 4

POLG_HCVTW STANDARD; PRT; 3010 AA.
 AC P29846;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN) (P22);
 DE ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35); ENVELOPE GLYCOPROTEIN E2
 DE (GP68) (GP70) (NS1); PROTEIN P7; NONSTRUCTURAL PROTEIN NS2 (P21)
 DE (EC 3.4.22.-); PROTEINASE/HELICASE NS3 (P70) (EC 3.4.21.-);
 DE NONSTRUCTURAL PROTEIN NS4A (P4); NONSTRUCTURAL PROTEIN NS4B (P27);
 DE NONSTRUCTURAL PROTEIN NS5A (P56); NONSTRUCTURAL PROTEIN NS5B (P66)
 DE (P70) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate Taiwan) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-92230206; PubMed-1314449;
 RA Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;
 RA "The Taiwanese hepatitis C virus genome: sequence determination and
 RA mapping the 5' terminus of viral genomic and antigenomic RNA.";
 RL Virology 188:102-113(1992).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE

HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
-1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND MRNA.
-1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: M84754; -; NOT_ANNOTATED_CDS.
PIR: A40244; GNWVTV.
HSP: P27958; 1HEI.
MEROPS: S29.001; -.
MEROPS: U39.001; -.
INTERPRO: IPR000745; -.
INTERPRO: IPR001490; -.
INTERPRO: IPR002166; -.
INTERPRO: IPR002518; -.
INTERPRO: IPR002519; -.
INTERPRO: IPR002521; -.
INTERPRO: IPR002522; -.
INTERPRO: IPR002531; -.
INTERPRO: IPR002868; -.
PFAM: PF01560; HCV_NS1; 1.
PFAM: PF01538; HCV_NS2; 1.
PFAM: PF01006; HCV_NS4a; 1.
PFAM: PF01001; HCV_NS4b; 1.
PFAM: PF01506; HCV_NS5a; 1.
PFAM: PF00998; HCV_RDRP; 1.
PFAM: PF01543; HCV_capsid; 1.
PFAM: PF01542; HCV_core; 1.
PFAM: PF01539; HCV_env; 1.
Polyprotein: Glycoprotein; Transferase; RNA-directed RNA polymerase; Core protein; Coat protein; Envelope protein; Helicase; ATP-binding; Transmembrane; Nonstructural
INIT_MET 1
CHAIN 1 115
CHAIN 116 191
CHAIN 192 383
CHAIN 384 729
CHAIN 730 1006
CHAIN 1007 1615
CHAIN 1616 1862
CHAIN 1863 2013
CHAIN 2014 3010
CHAIN TRANSMEM 347 369
ACT_SITE 1083 1083
ACT_SITE 1107 1107
ACT_SITE 1155 1165
SITE_BIND 1230 1337
SITE NP 1316 1319
CARBOHYD 196 196
CARBOHYD 209 209
CARBOHYD 233 233
CARBOHYD 234 234
CARBOHYD 250 250
CARBOHYD 305 305
CARBOHYD 417 417
CARBOHYD 423 423
CARBOHYD 430 430
CARBOHYD 448 448
CARBOHYD 532 532
CARBOHYD 540 540
CARBOHYD 556 556
CARBOHYD 576 576

FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 3010 AA; 327047 MW; AAD267D55CDFE215 CRC64;

Query Match 83.1%; Score 1710; DB 1; Length 3010;
Best Local Similarity 79.6%; Pred. No. 2.4e-137;
Matches 288; Conservative 33; Mismatches 41; Indels 0; Caps 0;

QY 2 THVTGSGAGHTVSGFVSLAPAGAKQNVQLINTNGSWHLNSTALNCNDSLNTGWLGLFYH 61
DB 385 TIVSGGTAVRTTHSLASLFTQASQKIQIINTNGSWHLNSTALNCNDSLNTGWLGLFYH 61
QY 62 HKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDPDPCYCHYPPKCGIYPAKSVCG 121
DB 445 HRFNASGCPERMASCRSIDKFDQGWGPITVTEADIODPCYCHYAPRPGIYPAKSVCG 504
QY 122 PVYCFTPSPVVVGTDRSGAPTYSWGENDTDVFLNNTPPPLGNWFCCTWMNSTGFTKVC 181
DB 505 PVYCFTPSPVVVGTDRFGAPTYSWGNETDVLILNTRPPQGNWFCCTWMNSTGFTKVC 564
QY 182 GAPPVIGGAGNNTLHCPTDCFRKHDPATYSRGGSGPWITPRCLVDYVYRLWHYPTINY 241
DB 565 GGPPCNIIGGGNNTLVCPDCEFRKHDPATYTCGSGPWLTPRCMVDPYVRLWHYPTVNF 624
QY 242 TIFKIRMYGVGVEHRLAECNNTNRGERCDLEDRSELSPLLLTTTQWVLPSCFTTLPA 301
DB 625 TIFKIRMYGVGVEHRLAECNNTNRGERCDLEDRSELSPLLLTTTQWVLPSCFTTLPA 684
QY 302 LSTGLIHLHQNVDVQVLYGVSSIASWAIKWVYVLLFLLADARVCSCLMMLLSOA 361
DB 685 LSTGLIHLHQNVDVQVLYGVSSIASWAIKWVYVLLFLLADARVCSCLMMLLSOA 744
QY 362 EA 363
DB 745 EA 746

RESULT 5
POLG-HCVJA STANDARD; PRT: 3010 AA.
ID POLG-HCVJA AC P26662;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN) (P22); ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35); ENVELOPE GLYCOPROTEIN E2 (GP68) (GP70) (NS1); PROTEIN P7; NONSTRUCTURAL PROTEIN NS2 (P21) (EC 3.4.22.-); PROTEIN P8; PROTEIN P9; NONSTRUCTURAL PROTEIN NS4 (P27); NONSTRUCTURAL PROTEIN NS4A (P4); NONSTRUCTURAL PROTEIN NS5 (P27); NONSTRUCTURAL PROTEIN NS5A (P56); NONSTRUCTURAL PROTEIN NS5B (P66) (P70) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate Japanese) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91088550; Pubmed=2175903;
RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S., Sugimura T., Shimotohno K.;
RT "Molecular cloning of the human hepatitis C virus genome from Japanese patients with non-A, non-B hepatitis";
RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
RN [2]
RP DISCUSSION OF SEQUENCE.
RX MEDLINE=91192160; Pubmed=1849488;
RA Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraliso K., Ohkoshi S., Shimotohno K.;

RT "Molecular structure of the Japanese hepatitis C viral genome.";
 RL FEBS Lett. 280:325-328(1991).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPID PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D90208; BAA14233.1; -;
 CC PTR; A39253; GNWVCJ.
 CC HSP; P27958; 1HEI.
 CC MEROPS; S29.001; -;
 CC INTERPRO; IPR000745; -;
 CC INTERPRO; IPR001490; -;
 CC INTERPRO; IPR002166; -;
 CC INTERPRO; IPR002518; -;
 CC INTERPRO; IPR002519; -;
 CC INTERPRO; IPR002521; -;
 CC INTERPRO; IPR002522; -;
 CC INTERPRO; IPR002531; -;
 CC INTERPRO; IPR002868; -;
 CC PFAM; PF01560; HCV_NS1; 1;
 CC PFAM; PF01538; HCV_NS2; 1;
 CC PFAM; PF01006; HCV_NS4; 1;
 CC PFAM; PF01001; HCV_NS4b; 1;
 CC PFAM; PF01506; HCV_NS5a; 1;
 CC PFAM; PF00998; HCV_RdRP; 1;
 CC PFAM; PF01543; HCV_capsid; 1;
 CC PFAM; PF01542; HCV_core; 1;
 CC PFAM; PF01539; HCV_env; 1;
 CC
 CC Polyprotein; Glycoprotein; Transferrase; RNA-directed RNA polymerase;
 CC Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 CC Transmembrane; Nonstructural
 CC INIT_MET 1 1
 CC CHAIN 1 115
 CC CHAIN 116 191
 CC CHAIN 192 383
 CC CHAIN 384 729
 CC CHAIN 730 1006
 CC CHAIN 1007 1615
 CC CHAIN 1616 1862
 CC CHAIN 1863 2013
 CC CHAIN 2014 3010
 CC TRANS 347 369
 CC ACT_SITE 1083 1083
 CC ACT_SITE 1107 1107
 CC ACT_SITE 1165 1165
 CC NP_BIND 1230 1237
 CC SITE 1316 1319
 CC CARBOHYD 196 196
 CC CARBOHYD 209 209
 CC CARBOHYD 234 234
 CC CARBOHYD 250 250
 CC CARBOHYD 305 305
 CC CARBOHYD 417 417
 CC CARBOHYD 423 423
 CC CARBOHYD 430 430
 CC CARBOHYD 448 448
 CC CARBOHYD 532 532
 CC CARBOHYD 556 556

FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 3010 AA; 327017 MW; AA993794F46DB185 CRC64;
 Query Match 82.9%; Score 1706; DB 1; Length 3010;
 Best Local Similarity 78.7%; Pred. No. 5.2e-137;
 Matches 285; Conservative 39; Mismatches 38; Indels 0; Gaps 0;
 QY 2 THVTGSGAGHTVSGFVSLAPGAKQNVQLINTNGSWHLNSTALNCNDSLNTGWLGLFYH 614
 DB 385 THVTGGRVASTQSLVSWLSQPSQKQLQVNTNGSWHLNRTALNCNDSLQGTGFAALEFYA 444
 QY 62 HKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGPDQRPYCHYPPKPCGIVPAKSVCG 121
 DB 445 HRFNASGCPERMASCRPTDEFAQGWGPIITHDMPSSDQRPYCHYAPRCPGIVPASQVCG 504
 QY 122 PVYCFTPSPVVVGTDRSGAPTYSGWENDTDVFLVNNTRPPLGNWFGCTWNNSTGFTKVC 181
 DB 505 PVYCFTPSPVVVGTDRFGAPTYSGWENETDVLNTRPPGQWFGCTWNNSTGFTKTC 564
 QY 182 GAPPVIGGAGNNTLHCPITDCFRKHDPATYSRCSGSPWITPRCLVDYPIYLWHYPCTINY 241
 DB 565 GGPECNIGGVGNNTLVCPITDCFRKHDPATYTKCGSPWLTFRCMVDYPIYLWHYPCTVNF 624
 QY 242 TIFKIRMYVGVGVEHRLAACAACNWTGRCERCDLDRDRSELSPLLLTTQWQVLPSCFTTLP 301
 DB 625 TVFKIRMYVGVGVEHRLAACAACNWTGRCERCDLDRDRSELSPLLLTTQWQVLPSCFTTLP 684
 QY 302 LSTGLIHLHQNIVDVQYLYGVGSSIASWAKWEYVLLFLLADARVCSCLWMLLSOA 361
 DB 685 LSTGLIHLHNRNIVDVQYLYGVGSSIASWAKWEYVLLFLLADARVCSCLWMLLSOA 744
 QY 362 EA 363
 DB 745 EA 746
 RESULT 6
 POLG_HCVJT
 ID POLG_HCVJT STANDARD; PRT; 3010 AA.
 AC Q00269;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN) (P22);
 DE ENVELOPE GLYCOPROTEIN E1 (GP32); ENVELOPE GLYCOPROTEIN E2
 DE (GP68) (GP70) (NS1); PROTEIN P7; NONSTRUCTURAL PROTEIN NS2 (P21)
 DE (SC 3.4.22.-); PROTEASE/HELICASE NS3 (P70) (EC 3.4.21.-);
 DE NONSTRUCTURAL PROTEIN NS4A (P4); NONSTRUCTURAL PROTEIN NS4B (P27);
 DE NONSTRUCTURAL PROTEIN NS5A (P56); NONSTRUCTURAL PROTEIN NS5B (P66)
 DE (P70) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-JT) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92295714; PubMed=1318627;
 RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,
 RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
 RT "Molecular cloning of hepatitis C virus genome from a single Japanese
 RT carrier: sequence variation within the same individual and among
 RT infected individuals.";
 RL Virus Res. 23:39-53(1992).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A

CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC
 CC 1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.
 CC -----

DR EMBL: D11168; BAA01943.1; -
 DR PIR: A45573; A45573.
 DR HSP: P27958; 1HEI.
 DR MEROPS: S29.001; -
 DR MEROPS: U39.001; -
 DR INTERPRO: IPR000745; -
 DR INTERPRO: IPR001490; -
 DR INTERPRO: IPR002166; -
 DR INTERPRO: IPR002518; -
 DR INTERPRO: IPR002519; -
 DR INTERPRO: IPR002521; -
 DR INTERPRO: IPR002522; -
 DR INTERPRO: IPR002531; -
 DR INTERPRO: IPR002868; -
 DR PFAM: PF01560; HCV_NS1; 1.
 DR PFAM: PF01538; HCV_NS2; 1.
 DR PFAM: PF01006; HCV_NS4a; 1.
 DR PFAM: PF01001; HCV_NS4b; 1.
 DR PFAM: PF01506; HCV_NS5a; 1.
 DR PFAM: PF00998; HCV_RDRP; 1.
 DR PFAM: PF01543; HCV_capsid; 1.
 DR PFAM: PF01542; HCV_env; 1.
 DR PFAM: PF01539; HCV_gag; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural
 FT INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT CHAIN 3010 369
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT ACT_SITE 1230 1230
 FT SITE 1316 1319
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 532 532
 FT CARBOHYD 540 540
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077

FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 3010 AA; 326573 MW; 94A1C77435D642BB CRC64;
 Query Match 81.2%; Score 1671; DB 1; Length 3010;
 Best Local Similarity 78.5%; Pred. No. 4.9e-134;
 Matches 284; Conservative 29; Mismatches 49; Indels 0; Gaps 0;
 QY 2 THVTGSGAGHTVSGFVSLAPGAKQNVQLINTNGSHLSTALNCNDSLNTGWLGLFHY 61
 DB 385 TVTTGSGQARHTQSVTFEFTQGPQRIQLINTNGSHLSTALNCNDSLNTGWLGLFHY 444
 QY 62 HKFNSSGCPERLASCRPLTDFOGWGPISYANGSGDPQRPYCHWHYPPKPGIYPAKSVCG 121
 DB 445 HKFNSSGCPERLASCRPLTDFOGWGPISYANGSGDPQRPYCHWHYPPKPGIYPAKSVCG 504
 QY 122 PVYCFTPSPVYVCTTDRSGAPTYSGENDTDVFLANTRPPLGNWFGCTWMNSTGFTKVC 181
 DB 505 PVYCFTPSPVYVCTTDRSGAPTYSGENDTDVFLANTRPPLGNWFGCTWMNSTGFTKVC 564
 QY 182 GAPPVIGGAGNNTLCPTDCERKHPDATYSRGSGPWITPRCLVDYPRYRLHYPCVTINY 241
 DB 565 GAPPVIGGAGNNTLCPTDCERKHPDATYSRGSGPWITPRCLVDYPRYRLHYPCVTINY 624
 QY 242 TIFKIRMYGGVEHRLLEAACNMTRGERCDLEDRSELSPLLLTTTOMOVLPSCFTTLPA 301
 DB 625 TIFKIRMYGGVEHRLLEAACNMTRGERCDLEDRSELSPLLLTTTOMOVLPSCFTTLPA 684
 QY 302 LSTGLHLHQNIVDVOYLYGVSSIASWAKWYVYVLLLELLADARVCCLNMLLSOA 361
 DB 685 LSTGLHLHQNIVDVOYLYGVSSIASWAKWYVYVLLLELLADARVCCLNMLLSOA 744
 QY 362 EA 363
 DB 745 EA 746
 RESULT 7
 POLG_HCVJ8 STANDARD; PRT: 3033 AA.
 AC P26661;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN) (P22);
 DE ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35); ENVELOPE GLYCOPROTEIN E2
 DE (GP68) (GP70) (NS1); PROTEIN P7; NONSTRUCTURAL PROTEIN NS2 (P21)
 DE (EC 3.4.22.-); PROTEINASE/HELICASE NS3 (P70) (EC 3.4.21.-);
 DE NONSTRUCTURAL PROTEIN NS4A (P4); NONSTRUCTURAL PROTEIN NS4B (P27);
 DE NONSTRUCTURAL PROTEIN NS5A (P56); NONSTRUCTURAL PROTEIN NS5B (P66)
 DE (P70) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-J8) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-92230232; PubMed-1314459;
 RA Okamoto H., Kurai K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T.,
 RA Fukuda S., Tsuda F., Mishiro S.;
 RT "Full-length sequence of a hepatitis C virus genome having poor
 RT homology to reported isolates: comparative study of four distinct
 RT genotypes";
 RL Virology 188:331-341(1992).
 CC 1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC 1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC 1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

```

CC EMBL: D00944; BAA00792.1; -
CC PIR: JQ1303; JQ1303.
CC HSP: P27958; 1HE1.
CC MEROPS: S29.001; -
CC MEROPS: U39.001; -
CC INTERPRO: IPR000745; -
CC INTERPRO: IPR001490; -
CC INTERPRO: IPR002166; -
CC INTERPRO: IPR002518; -
CC INTERPRO: IPR002519; -
CC INTERPRO: IPR002521; -
CC INTERPRO: IPR002522; -
CC INTERPRO: IPR002531; -
CC INTERPRO: IPR002868; -
CC PFAM: PF01560; HCV_NS1; 1.
CC PFAM: PF01538; HCV_NS2; 1.
CC PFAM: PF01006; HCV_NS4a; 1.
CC PFAM: PF01001; HCV_NS4b; 1.
CC PFAM: PF01506; HCV_NS5a; 1.
CC PFAM: PF00998; HCV_RDRP; 1.
CC PFAM: PF01543; HCV_capsid; 1.
CC PFAM: PF01542; HCV_core; 1.
CC PFAM: PF01539; HCV_env; 1.
CC PFAM: PF01539; HCV_env; 1.
CC KW Polyprotein: Glycoprotein; Transferase: RNA-directed RNA polymerase;
CC KW Core protein: Coat protein; Helicase; ATP-binding;
CC KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
CC FT INIT_MET 1 1
CC FT CHAIN 1 115
CC FT CHAIN 116 191
CC FT CHAIN 192 383
CC FT CHAIN 384 733
CC FT CHAIN 734 1010
CC FT CHAIN 1011 1619
CC FT CHAIN 1620 1866
CC FT CHAIN 1867 2017
CC FT CHAIN 2018 3033
CC FT CHAIN 3033 369
CC FT TRANSMEM 347 369
CC FT ACT_SITE 1087 1087
CC FT ACT_SITE 1111 1111
CC FT ACT_SITE 1169 1169
CC FT NP_BIND 1234 1241
CC FT SITE 1320 1323
CC FT CARBOHYD 196 196
CC FT CARBOHYD 209 209
CC FT CARBOHYD 234 234
CC FT CARBOHYD 305 305
CC FT CARBOHYD 417 417
CC FT CARBOHYD 423 423
CC FT CARBOHYD 430 430
CC FT CARBOHYD 448 448
CC FT CARBOHYD 477 477
CC FT CARBOHYD 534 534
CC FT CARBOHYD 542 542
CC FT CARBOHYD 558 558
CC FT CARBOHYD 578 578
CC FT CARBOHYD 627 627
CC FT CARBOHYD 649 649
CC FT CARBOHYD 1091 1091
CC FT CARBOHYD 2038 2038
CC FT CARBOHYD 2811 2811
CC FT CARBOHYD 3033 3033
CC FT SEQUENCE 3033 AA; 329165 MW; F957F5C1A273B89E CRC64;

use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).
```

Query Match

Best Local Similarity 74.4%; Score 1531; DB 1; Length 3033;

Matches 258; Conservative 51; Mismatches 54; Indels 4; Gaps 3;

```

QY 1 ETHVGTGSAGHTVSGFVSLAPGAKQKVQLINTNGSHLSTALNCNDSLNTGWLACLFY 60
DB 384 QTHVGTGSAGHTVSGFVSLAPGAKQKVQLINTNGSHLSTALNCNDSLNTGWLACLFY 443
QY 61 HHKFNSSGCCPERLASCRPLTDFDQGCPISEA-NGSCP-DQRPYCHYHPKPGCIVPAKS 118
DB 444 THSFNSSGCCPERLASCRPLTDFDQGCPISEA-NGSCP-DQRPYCHYHPKPGCIVPAKS 503
QY 119 VCGPVYCFPTSPVVVGTTRDSCGAPTYSWGENDTDFVFLNNTTRPLGNWFCCTWMNSTGET 178
DB 504 VCGPVYCFPTSPVVVGTTRDSCGAPTYSWGENDTDFVFLNNTTRPLGNWFCCTWMNSTGET 563
QY 179 KVGAPPCVITGGAGNNT--LHCPTDCEFRKHPDATYKRCGSGPWITPRCLVDYYPYRLWHYP 236
DB 564 KTCGAPPCVITGGAGNNT--LHCPTDCEFRKHPDATYKRCGSGPWITPRCLVDYYPYRLWHYP 623
QY 237 CTINYTIKIRMYVGVGVEHRLAECANWTRGCDLEDRDRSELSPILLTTOHQVLPSCF 296
DB 624 CTINYTIKIRMYVGVGVEHRLAECANWTRGCDLEDRDRSELSPILLTTOHQVLPSCF 683
QY 297 TTLPALSTGLIHLHQNIVDQYLYGVGSSIASMAIKWEYVVLFLLLADARVCSCLMML 356
DB 684 SDLPALSTGLIHLHQNIVDQYLYGVGSSIASMAIKWEYVVLFLLLADARVCSCLMML 743
QY 357 LISQAEA 363
DB 744 LLGQAEA 750

RESULT 9
POLG_HCVJ7 STANDARD; PRT; 737 AA.
AC P27961;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN); MATRIX
DE PROTEIN (ENVELOPE PROTEIN M); MAJOR ENVELOPE PROTEIN E; NONSTRUCTURAL
DE PROTEINS NS1 AND NS2] (FRAGMENT).
OS Hepatitis C virus (isolate HC-J7) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-92230232; PubMed-1314459;
RA Okamoto H., Kura K., Okada S.I., Yamamoto K., Lizuka H.,
RA Tanaka T., Fukuda S., Tanaka F., Mishiro S.;
RT "Full-length sequence of a hepatitis C virus genome having poor
RT homology to reported isolates: comparative study of four distinct
RT genotypes.";
RL Virology 188:331-341(1992).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
```


DR PFAM; PF01543; HCV_capsid; 1.
DR PFAM; PF01542; HCV_core; 1.
DR PFAM; PF01539; HCV_env; 1.
KW Polypeptide; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural protein.
FT INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 733
FT CHAIN 734 >737
FT TRANSMEM 347 369
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 233 233
FT CARBOHYD 299 299
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 477 477
FT CARBOHYD 534 534
FT CARBOHYD 542 542
FT CARBOHYD 558 558
FT CARBOHYD 578 578
FT CARBOHYD 627 627
FT CARBOHYD 649 649
FT NON_TER 737
SQ SEQUENCE 737 AA; 81691 MW; 67DFAE11854122F2 CRC64;

Query Match 73.3%; Score 1508; DB 1; Length 737;
Best Local Similarity 73.4%; Pred. No. 7.4e-121;
Matches 259; Conservative 37; Mismatches 53; Indels 4; Gaps 2;

QY 2 THVTGSGAGTGVGVVSLIAPGAKONVOLINTNGSHWLNSTALNCNDSLNTGWLGLFYH 61
DB 385 TQVTGGAHTVGVVSLIAPGAKONVOLINTNGSHWLNSTALNCNDSLNTGWLGLFYH 444
QY 62 HFNSSGCPERLASRPLTDFQGGPISYANG--SGDPQRYCHYPPKGIYPAKSV 119
DB 445 HFNSSGCPERLASRPLTDFQGGPISYANG--SGDPQRYCHYPPKGIYPAKSV 504
QY 120 CGPVCFPTSPVVGTTDRSGAPTSYANG--SGDPQRYCHYPPKGIYPAKSV 179
DB 505 CGPVCFPTSPVVGTTDRSGAPTSYANG--SGDPQRYCHYPPKGIYPAKSV 564
QY 180 VCGAPPCTVGGAGNTLH--CPTDCFRKHPDATYSCSGGPPWITPRCLVDYPRYLWHYPC 237
DB 565 TCGAPPCTVGGAGNTLH--CPTDCFRKHPDATYSCSGGPPWITPRCLVDYPRYLWHYPC 624
QY 238 TNYIFKRMVYGVGVEHLEACNWTGRCEDLDRBSELSPLLLTTOQVLPSCFT 297
DB 625 TNYIFKRMVYGVGVEHLEACNWTGRCEDLDRBSELSPLLLTTOQVLPSCFT 684
QY 298 TLPLSTGLHLHQNIVDVQVLYGVGSSSTASWAKWYVWVLLFLIADARVCS 350
DB 685 DLPLSTGLHLHQNIVDVQVLYGVGSSSTASWAKWYVWVLLFLIADARVCS 737

RESULT 10
POLG_HCVJ5
ID POLG_HCVJ5 STANDARD; PRT; 737 AA.
AC 27960;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN); MATRIX
DE PROTEINS NS1 AND NS2] (FRAGMENT).
OS Hepatitis C virus (isolate HC-J5) (HCV).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
RN Hepacivirus.
RN SEQUENCE FROM N.A.
RX MEDLINE=92230232; PubMed=1314459;
RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
RT "Full-length sequence of a hepatitis C virus genome having poor
homology to reported isolates: comparative study of four distinct
genotypes.";
RT Virology 188:331-341(1992).
RL -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed, usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; D10075; BAA00969.1;
DR INTERPRO; IPR002519;
DR INTERPRO; IPR002521;
DR INTERPRO; IPR002522;
DR INTERPRO; IPR002531;
DR PFAM; PF01560; HCV_NS1; 1.
DR PFAM; PF01543; HCV_capsid; 1.
DR PFAM; PF01542; HCV_core; 1.
DR PFAM; PF01539; HCV_env; 1.
KW Polypeptide; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural protein.
KW INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 733
FT CHAIN 734 >737
FT TRANSMEM 347 369
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 234 234
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 477 477
FT CARBOHYD 534 534
FT CARBOHYD 542 542
FT CARBOHYD 558 558
FT CARBOHYD 578 578
FT CARBOHYD 627 627
FT CARBOHYD 649 649
FT NON_TER 737
SQ SEQUENCE 737 AA; 81207 MW; 3AF699D82AD501B1 CRC64;

Query Match 71.6%; Score 1473; DB 1; Length 737;
Best Local Similarity 71.4%; Pred. No. 7.1e-118;
Matches 252; Conservative 45; Mismatches 52; Indels 4; Gaps 3;

QY 2 THVTGSGAGTGVGVVSLIAPGAKONVOLINTNGSHWLNSTALNCNDSLNTGWLGLFYH 61
DB 385 TQVTGGAHTVGVVSLIAPGAKONVOLINTNGSHWLNSTALNCNDSLNTGWLGLFYH 444


```
Qy 62 HKFNSSGCPERLASRPLTDFDQGWGPISEA-NGSGP-DQRPYCHWHYPPKCGIYPAKSV 119
Db 445 NRENSGCPHRLSVCRSEAFRIWGTLOEDNVNTPEDMRPYCHWHYPPKCGIYPAKSV 504
Qy 120 CGPVYCTPSPVVGTTDRSGAPTYSGENDTVFLNNTNRPPLGNWFGCTWMMNSTGFTK 179
Db 505 CGPVYCTPSPVVGTTDARGVPTTYWGENETDVELLNSTRPRGSGWFGCTWMMNSTGFTK 564
Qy 180 VCGAPPVIGGAGN-NWTLCHPTDCFRKHPDATYRCGSGPWITPRCLVDYPYRLWHPYC 237
Db 565 TCGAPPCKIRADFNASTDLLCTDCFRKHSDATYIKCGSGFWLTPKCMVDYPYRLWHPYC 624
Qy 238 TINTYFKIRMYVGVEHRLAAACNWTGRCERCDLDRSRLSPLLTTLTQWQVLPSCFT 297
Db 625 TVNYSIEKIRMYVGVEHRLTAACNFTRGDCNLEDRDRSRLSPLLTTLTQWQVLPSCFT 684
Qy 298 TLPALSTGLHLHONIVDVQYLVGVGSSIASWATKWEYVVLFLLLADARVCS 350
Db 685 DLPALSTGLHLHONIVDVQYLVGVGSSIASWATKWEYVVLFLLLADARVCA 737

RESULT 11
POLG_HCVH4
ID POLG_HCVH4 STANDARD: PRT: 520 AA.
AC Q01404;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN) (P22);
DE ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35); ENVELOPE GLYCOPROTEIN E2
DE (GP68) (GP70) (NS1)] (FRAGMENT).
OS Hepatitis C virus (isolate HCV-476) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93019030; PubMed=1383400;
RA Abe K., Inchauspe G., Fujisawa K.;
RT "Genomic characterization and mutation rate of hepatitis C virus
RT isolated from a patient who contracted hepatitis during an epidemic
RT of non-A, non-B hepatitis in Japan.";
RL J. Gen. Virol. 73:2725-2729(1992).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D10688; BAA01530.1;
CC DR INTERPRO: IPR002519;
CC DR INTERPRO: IPR002521;
CC DR INTERPRO: IPR002522;
CC DR INTERPRO: IPR002531;
CC DR PFAM: PF01560; HCV_Ns1; 1;
CC DR PFAM: PF01543; HCV_capsid; 1;
CC DR PFAM: PF01542; HCV_core; 1;
CC DR PFAM: PF01539; HCV_env; 1;
CC DR Polyprotein; Glycoprotein; Coat protein; Envelope protein;
CC Transmembrane; Nonstructural protein.
CC REMOVED FROM CAPSID PROTEIN C BY THE
CC INIT_MET 1
CC CELLULAR AMINOPEPTIDASE.
CC CHAIN 1 115
CC CHAIN 116 191
CC CHAIN 192 383
CC CHAIN 384 >520
CC CHAIN 520 347
CC TRANSMEM 347
```

```
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 418 418 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 424 424 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 431 431 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 449 449 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 520 520 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 520 AA; 56499 MW; AAL35246CF20D525 CRC64;

Query Match 27.0%; Score 555; DB 1; Length 520;
Best Local Similarity 68.9%; Pred. No. 6.2e-40;
Matches 93; Conservative 12; Mismatches 30; Indels 0; Gaps 0;

QY 2 THVTGSGAGTTSVGFYSLAPGAKQNVOLINTNGSHLNSALNCNDSLNTGWLGLFYH 61
DB 386 THVTGSGAGTTSVGFYSLAPGAKQNVOLINTNGSHLNSALNCNDSLNTGWLGLFYH 61
QY 62 HKFNSSGCPERLASRPLTDFDQGWGPISEA-NGSGP-DQRPYCHWHYPPKCGIYPAKSV 121
DB 446 HKFNSSGCPERLASRPLTDFDQGWGPISEA-NGSGP-DQRPYCHWHYPPKCGIYPAKSV 121
QY 122 PVYCTPSPVVGTT 136
DB 506 PVYCTPSPVVGTT 520

RESULT 12
POLG_HCVJ2
ID POLG_HCVJ2 STANDARD: PRT: 513 AA.
AC P27959;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN) (P22);
DE ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35); ENVELOPE GLYCOPROTEIN E2
DE (GP68) (GP70) (NS1)] (FRAGMENT).
OS Hepatitis C virus (isolate HC-J2) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92230232; PubMed=1314459;
RA Okamoto H., Kuri K., Okada S.I., Yamamoto K., Lizuka H.,
RA Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
RT "Full-length sequence of a hepatitis C virus genome having poor
RT homology to reported isolates: comparative study of four distinct
RT genotypes.";
RL Virology 188:331-341(1992).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D10074; BAA00568.1;
CC DR INTERPRO: IPR002519;
CC DR INTERPRO: IPR002521;
CC DR INTERPRO: IPR002522;
CC DR INTERPRO: IPR002531;
CC DR PFAM: PF01560; HCV_Ns1; 1;
CC PFAM: PF01543; HCV_capsid; 1;
CC PFAM: PF01542; HCV_core; 1;
CC Polyprotein; Glycoprotein; Coat protein; Envelope protein;
CC Transmembrane; Nonstructural protein.
CC REMOVED FROM CAPSID PROTEIN C BY THE
CC INIT_MET 1
CC CELLULAR AMINOPEPTIDASE.
CC CHAIN 1 115
CC CHAIN 116 191
CC CHAIN 192 383
CC CHAIN 384 >520
CC CHAIN 520 347
CC TRANSMEM 347
```

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

DR	PFAM; PF01543; HCV_capsid; 1.
DR	PFAM; PF01542; HCV_core; 1.
DR	PFAM; PF01539; HCV_env; 1.
KW	Polyprotein; Glycoprotein; Coat protein; Envelope protein;
FT	Transmembrane; Nonstructural protein; Removed from capsid protein C by the INIT_MET 1 1
FT	CELLULAR AMINOPEPTIDASE.
FT	CAPSID PROTEIN C (POTENTIAL).
FT	MATRIX PROTEIN (POTENTIAL).
FT	MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT	NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).
FT	POTENTIAL.
FT	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 196 196
FT	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 209 209
FT	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 233 233
FT	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 234 234
FT	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 250 250
FT	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 305 305
FT	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 417 417
FT	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 423 423
FT	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 430 430
FT	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 448 448
FT	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	NON_TER 513 513
SO	SEQUENCE 513 AA; 55704 MW; 943F31E3514CDEF3 CRC64;

	Query Match	26.98;	Score 553;	DB 1;	Length 513;
	Best Local Similarity	72.1%;	Pred. No.	9e-40;	
	Matches	93;	Mismatches	26;	Indels
		Conservative	0;	Gaps	0;
Qy	2	THVTGGSAGHTVSGFVSLLAPGAKNVOLINTNGSWHLNSTALNCNDSLNTGWLAGLFYH	61		
Dd	385	THVTGGATGHHTTGTAISFLPGASOKIOLINTNGSWHLNRPTALNCNDSLNTGFFLAALFYT	444		
Qy	62	HKFNSGCGPERLASCRPLTDQQWGPISTANGSFGDQRPCWHYPKPCGIIVPAKSVC	121		
Dd	445	HKFNASGCGPERLASCSRIDGFDQGWPITYTEPGSDSKPKCWHYAQCRCSVSAADVCG	504		
Qy	122	PVYCFTTPSP	130		
Dd	505	PVYCFTTPSP	513		

```

RESULT 13
ID POLG_HCVHK STANDARD: PRT; 520 AA.
AC Q01403;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN) (P22);
DE ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35); ENVELOPE GLYCOPROTEIN E2
DE (GP68) (GP70) (NS1)] (FRAGMENT).
OS Hepatitis C virus (isolate HCV-KF) (HCV).
CC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
CC Hepacivirus.
CC [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=93019030; PubMed=1383400;
RX Abe K., Inchauspe G., Fujisawa K.;
RA "Genomic characterization and mutation rate of hepatitis C virus
RT isolated from a patient who contracted hepatitis during an epidemic
RT of non-A, non-B hepatitis in Japan.";
RL J. Gen. Virol. 73:2725-2729(1992).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.

```

CC	modified and this statement is not removed.	usage by and for commo
CC	entitles requires a license agreement (See http://www.isb-sib.ch/annou	
CC	or send an email to license@isb-sib.ch).	
CC	-----	
DR	EMBL; DI0687; BAA01529.1; -	
DR	PIR; JQ1925; JQ1925.	
DR	INTERPRO; IPR002519; -	
DR	INTERPRO; IPR002521; -	
DR	INTERPRO; IPR002522; -	
DR	INTERPRO; IPR002531; -	
DR	PFAM; PF01560; HCV_NSI; 1.	
DR	PFAM; PF01543; HCV_capsid; 1.	
DR	PFAM; PF01342; HCV_core; 1.	
DR	PFAM; PF01539; HCV_env; 1.	
KW	Polyprotein; Glycoprotein; Coat protein; Envelope protein;	
KW	Transmembrane; Nonstructural protein.	
FT	INIT_MET	1
FT		1
FT	CHAIN	1
FT	CHAIN	115
FT	CHAIN	116
FT	CHAIN	131
FT	CHAIN	192
FT	CHAIN	384
FT	CHAIN	>520
FT	TRANSMEM	347
FT	CARBOHYD	369
FT	CARBOHYD	196
FT	CARBOHYD	196
FT	CARBOHYD	209
FT	CARBOHYD	233
FT	CARBOHYD	234
FT	CARBOHYD	234
FT	CARBOHYD	305
FT	CARBOHYD	305
FT	CARBOHYD	418
FT	CARBOHYD	418
FT	CARBOHYD	424
FT	CARBOHYD	424
FT	CARBOHYD	431
FT	CARBOHYD	431
FT	CARBOHYD	449
FT	CARBOHYD	449
FT	NON_TER	520
FT	NON_TER	520
SQ	SEQUENCE	520 AA; 56476 MW; 1D2BD0A6FF27349B CRC64;

```

Query Match          26.68; Score 548; DB 1; Length 520;
Best Local Similarity 68.1%; Pred. No. 2.4e-39;
Matches 92; Conservative 12; Mismatches 31; Indels 0; Gaps

Qy 2  THVTGGAGHTVSGFVSLAPGAKONVOLINTNGSWHLNSTALNCNDSLNTGWLAGLFYH 61
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 386 THTVGGTEGATQRLTSLFALGPSQKIQLINTNGSWHINRTALNCNDSFKTGFALAEFY 445

Qy 62 HKFNSSGCPERLASCRPLTDFDQGWGPTSYANGSPDORPCWYHPKPGIGIVPAKSVCG 121
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 446 HKFNASGCPHEWASCRPIKDFDQGWGPTVYAEPSISEQRPCWYHPKPGITIPASEVCG 505

Qy 122 PVICFTSPVVVGGTT 136
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 506 PVICFTSPVVVGGTT 520

RESULT 14
POLG_HCVH8          STANDARD;          PRT;          321 AA.
AC  P27956;
DT  01-AUG-1992 (Rel. 23, Created)
DT  01-AUG-1992 (Rel. 23, Last sequence update)
DT  15-JUL-1999 (Rel. 38, Last annotation update)
DE  GENOME POLYPROTEIN [CONTAINS: MATRIX PROTEIN (ENVELOPE PROTEIN M);
DE  MAJOR ENVELOPE PROTEIN E; NONSTRUCTURAL PROTEIN NS1 (FRAGMENT).
OS  Hepatitis C virus (isolate Hc18) (HCV).
OC  Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC  Hepacivirus.
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=91112009; PubMed=1846505;
RA  Welner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J.,
RA  Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M.,
RA  Han J.H.;
RT  "Variable and hypervariable domains are found in the regions of HCV
RT  corresponding to the flavivirus envelope and NS1 proteins and the

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 11:54:53 ; Search time 75.33 Seconds

(without alignments)

564,801 Million cell updates/sec

Title: US-09-407-430-2

Perfect score: 2058

Sequence: 1 ETHVTGSGAGTSGFVSL...ADARVCSCLWMLLSQAEA 363

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_15:*

- 1: sp-archaea:*
- 2: sp-bacteria:*
- 3: sp-fungi:*
- 4: sp-human:*
- 5: sp-invertebrate:*
- 6: sp-mammal:*
- 7: sp-mhc:*
- 8: sp-organelle:*
- 9: sp-phage:*
- 10: sp-plant:*
- 11: sp-rodent:*
- 12: sp-virus:*
- 13: sp-vertebrate:*
- 14: sp-unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2046	99.4	3011	12 Q91FE5	Q91FE5 hepatitis c
2	1957	95.1	778	12 Q04184	Q04184 h genome po
3	1949	94.7	3011	12 Q36579	Q36579 hepatitis c
4	1946	94.6	3011	12 Q36608	Q36608 hepatitis c
5	1943	94.4	3011	12 Q03463	Q03463 hepatitis c
6	1942	94.4	3011	12 Q36610	Q36610 hepatitis c
7	1932	93.9	3011	12 Q36609	Q36609 hepatitis c
8	1925	93.5	778	12 Q04185	Q04185 h genome po
9	1817	88.3	403	12 Q9PX22	Q9PX22 hepatitis c
10	1794	87.2	3010	12 Q9J3H0	Q9J3H0 hepatitis c
11	1790	87.0	3011	12 Q81754	Q81754 hepatitis c
12	1745	84.8	3010	12 P90194	P90194 hepatitis c
13	1744	84.7	3010	12 Q81760	Q81760 hepatitis c
14	1744	84.7	3010	12 Q9J3H9	Q9J3H9 hepatitis c
15	1742	84.6	3010	12 Q02828	Q02828 h genome po
16	1742	84.6	3010	12 Q9J3G1	Q9J3G1 hepatitis c
17	1741	84.6	2864	12 Q92975	Q92975 hepatitis c
18	1741	84.6	3010	12 Q9J3G8	Q9J3G8 hepatitis c
19	1740	84.5	2864	12 Q9WLK8	Q9WLK8 hepatitis c

20	1740	84.5	3010	12	Q92969	hepatitis c
21	1740	84.5	3010	12	Q81989	hepatitis c
22	1740	84.5	3013	12	Q9J3H4	hepatitis c
23	1739	84.5	3010	12	Q02829	h genome po
24	1738	84.5	3010	12	Q9Q1V6	hepatitis c
25	1737	84.4	2864	12	Q9WLLO	hepatitis c
26	1737	84.4	3010	12	Q92970	hepatitis c
27	1737	84.4	3010	12	Q93016	hepatitis c
28	1736	84.4	3010	12	Q9J3G4	hepatitis c
29	1735	84.3	3010	12	Q9J310	hepatitis c
30	1733	84.2	2864	12	Q92976	hepatitis c
31	1732	84.2	3015	12	Q9WPH5	hepatitis c
32	1731	84.1	3010	12	Q9Q1X4	hepatitis c
33	1731	84.1	3010	12	Q9Q1X2	hepatitis c
34	1731	84.1	3010	12	Q9J3H8	hepatitis c
35	1730	84.1	3010	12	Q9Q1X5	hepatitis c
36	1730	84.1	3010	12	Q9J3C2	hepatitis c
37	1729	84.0	3010	12	Q9J3C9	hepatitis c
38	1728	84.0	802	12	Q68520	hepatitis c
39	1728	84.0	2864	12	Q9WLK9	hepatitis c
40	1728	84.0	3010	12	Q92971	hepatitis c
41	1728	84.0	3010	12	P90192	hepatitis c
42	1728	84.0	3010	12	Q9Q6P1	hepatitis c
43	1728	84.0	3010	12	Q9J3H1	hepatitis c
44	1727	83.9	2864	12	Q92974	hepatitis c
45	1727	83.9	3010	12	Q9J3H3	hepatitis c

ALIGNMENTS

RESULT 1

Q91FE5
ID Q91FE5 PRELIMINARY; PRT; 3011 AA.
AC Q91FE5;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE POLYPROTEIN.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Lanford R.E., Lee H., Chavez D., Guerra B., Brasky K.;
RT "An Infectious Clone of the HCV-1 Prototype Sequence";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF271632; AAF81759.1;
SQ SEQUENCE 3011 AA; 327126 MW; 2489CE74AC864E58 CRC64;

Query Match	99.4%	Score	2046;	DB	12;	Length	3011;
Best Local Similarity	99.4%	Pred. No.	9.8e-189;				
Matches	361;	Conservative	0;	Mismatches	2;	Indels	0;
Gaps	0;						
Qy	1	ETHVTGSGAGTSGFVSL	APGAKQNVOLINTNGSWHLNSTALNCNDSLTGWLGLFY	60			
Db	384	ETHVTGSGAGTSGFVSL	APGAKQNVOLINTNGSWHLNSTALNCNDSLTGWLGLFY	443			
Qy	61	HKFNSSCCPERLASCRL	PTDQGWGPISYANGSGDPDPCYCHYPPKPGIYPAKSV	120			
Db	444	HKFNSSCCPERLASCRL	PTDQGWGPISYANGSGDPDPCYCHYPPKPGIYPAKSV	503			
Qy	121	GPVYCFPTSPVVVGT	TDPSGAPTSYWGENDTDVFLNNTRPPLGNWFCCTWMNSTGFTKV	180			
Db	504	GPVYCFPTSPVVVGT	TDPSGAPTSYWGENDTDVFLNNTRPPLGNWFCCTWMNSTGFTKV	563			
Qy	181	CGAPPCVIGGAGNNT	LHCPTDCFRKHDPDATYSRGSGPWTTPRCCLVDYPRKLWHYPCTIN	240			
Db	564	CGAPPCVIGGAGNNT	LHCPTDCFRKHDPDATYSRGSGPWTTPRCCLVDYPRKLWHYPCTIN	623			
Qy	241	YTIKIRMYGVGVEH	LEAACNWTGRGCDLEDRSLSPLLLTTTQWQVLPSCFTTLP	300			

```

Db 624 YTFKRMVYGVGVEHLEACNWTGRGCDLEDRSELSPLLLTTQWQVLPSCFTTLP 683
QY 301 ALSTGLHLHQNIVDYQVLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLMMLLSQ 360
Db 684 ALSTGLHLHQNIVDYQVLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLMMLLSQ 743
QY 361 AEA 363
Db 744 AEA 746

RESULT 2
Q04184 PRELIMINARY; PRT; 778 AA.
AC Q04184; Q81810;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C
(CORE PROTEIN); MATRIX PROTEIN
(ENVELOPE PROTEIN M); MAJOR ENVELOPE PROTEIN E; NONSTRUCTURAL PROTEIN
NS1; NONSTRUCTURAL PROTEIN NS2; NONSTRUCTURAL PROTEIN NS4A;
NONSTRUCTURAL PROTEIN NS4B; HELICASE
(NS3); RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (NS5)] (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91195357; PubMed=1849654;
RA Ozata N., Alter H.J., Miller R.H., Purcell R.H.;
RT "Nucleotide sequence and mutation rate of the H strain of hepatitis C
virus.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:3392-3396(1991).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION. NS3
AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA.
CC EMBL; M62381; AAB02127.1; -.
DR INTERPRO; IPR002518; -.
DR INTERPRO; IPR002519; -.
DR INTERPRO; IPR002521; -.
DR INTERPRO; IPR002522; -.
DR INTERPRO; IPR002523; -.
DR PFAM; PF01538; HCV_NS2; 1.
DR PFAM; PF01539; HCV_env; 1.
DR PFAM; PF01542; HCV_core; 1.
DR PFAM; PF01543; HCV_capsid; 1.
DR PFAM; PF01560; HCV_NS1; 1.
KW Polyprotein; Glycoprotein; RNA-directed RNA polymerase; Core protein;
Coat protein; Envelope protein; Helicase; ATP-binding; Transmembrane;
Nonstructural protein.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
CELLULAR AMINOPEPTIDASE.
FT CHAIN 2 115 CAPSID PROTEIN C (POTENTIAL).
FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2
(POTENTIAL).
FT CHAIN 730 >778 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT TRANSMEM 347 369 POTENTIAL.
FT NON_TER 778 778
SQ SEQUENCE 778 AA; 85143 MW; F7709172CD03E39B CRC64;

```

Query Match 95.1%; Score 1957; DB 12; Length 778;
 Best Local Similarity 94.2%; Pred. No. 8.3e-181;
 Matches 342; Conservative 10; Mismatches 11; Indels 0; Gaps 0;

```

QY 1 ETHVTGGSAGHTVSGFVSLAPGAKQNVOLINTNGSWHLNSTALNCNDSINTGWLGLFY 60
Db 384 ETHVTGGSAGRTTAGLVGLLTPGAKQNIQLINTNGSWHLNSTALNCNDSINTGWLGLFY 443
QY 61 HHKFNSSGCPERLASCRPLTDFDQGGPISYANGSPDQRPYCHWYPPKPCGIVPAKSVC 120
Db 444 HHKFNSSGCPERLASCRRLTDFDQGGPISYANGSGLDERPYCHWYPPKPCGIVPAKSVC 503
QY 121 GPVYCFTSPVGVVGTDRSGAPTSYWGENDTDFVLNNTRPPLGNWFGCTWMNSTGFTKV 180
Db 504 GPVYCFTSPVGVVGTDRSGAPTSYWGANDTDFVLNNTRPPLGNWFGCTWMNSTGFTKV 563
QY 181 CGAPPCVIGGAGNNTLHCPTDCFRKHPDATYSRCGSGPWITPRCLVDYDYPYRLWHYPCIN 240
Db 564 CGAPPCVIGGVGNNTLLCPTDCFRKHPDATYSRCGSGPWITPRCMVDYDYPYRLWHYPCIN 623
QY 241 YTFKRMVYGVGVEHLEACNWTGRGCDLEDRSELSPLLLTTQWQVLPSCFTTLP 300
Db 624 YTFKRMVYGVGVEHLEACNWTGRGCDLEDRSELSPLLLTTQWQVLPSCFTTLP 683
QY 301 ALSTGLHLHQNIVDYQVLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLMMLLSQ 360
Db 684 ALSTGLHLHQNIVDYQVLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLMMLLSQ 743
QY 361 AEA 363
Db 744 AEA 746

RESULT 3
O36579 PRELIMINARY; PRT; 3011 AA.
AC O36579;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE POLYPROTEIN.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=H77;
RX MEDLINE=97373636; PubMed=9228008;
RA Kolykhalov A.A., Agapov E.V., Blight K.J., Mihalik K., Feinstone S.M.,
Rice C.M.;
RT "Transmission of hepatitis C by intrahepatic inoculation with
transcribed RNA.";
RL Science 277:570-574(1997).
DR EMBL; AF009606; AAB66324.1; -.
DR INTERPRO; IPR000745; -.
DR INTERPRO; IPR001410; -.
DR INTERPRO; IPR001490; -.
DR INTERPRO; IPR002166; -.
DR INTERPRO; IPR002518; -.
DR INTERPRO; IPR002519; -.
DR INTERPRO; IPR002521; -.
DR INTERPRO; IPR002522; -.
DR INTERPRO; IPR002531; -.
DR INTERPRO; IPR002868; -.
DR PFAM; PF00998; HCV_RdRP; 1.
DR PFAM; PF01001; HCV_NS4b; 1.
DR PFAM; PF01006; HCV_NS4a; 1.
DR PFAM; PF01506; HCV_NS5a; 1.
DR PFAM; PF01538; HCV_NS2; 1.
DR PFAM; PF01539; HCV_env; 1.
DR PFAM; PF01542; HCV_core; 1.
DR PFAM; PF01543; HCV_capsid; 1.
DR PFAM; PF01560; HCV_NS1; 1.
DR PRODOM; PD186062; -. 1.
KW Polyprotein.

```

SQ SEQUENCE 3011 AA; 327184 MW; E2E0EE809C63C1B9 CRC64;

Query Match	94.7%;	Score 1949;	DB 12;	Length 3011;
Best Local Similarity	93.9%;	Pred. No. 2.3e-179;		
Matches 341;	Conservative 10;	Mismatches 12;	Indels 0;	

Qy	1	ETHWTGSAGHTVSGFVSVLLAPGAKQNVQLINTNGSWHLNSTALNCNDSLNTGWLAGLFY	60
Db	384	ETHWTGSAGRTTAGVLGLLTFGAKQNLQLTINTNGSWHLNSTALNCNDSLNTGWLAGLFY	443
Qy	61	HKFNSSGCCPERLASCRPLTDFDQGWGPISYANGSGDPDRPYCHWYPKPCGIYPAKSVC	120
Db	444	QHKFNSSGCCPERLASCRRLTDFAGWGPISYANGSGDLDERPYCHWYPPRPGCIYPAKSVC	503
Qy	121	GPVYCFTPSPVVVGTTRDSGAPTVSWGENDTDVFLNNTRPPLGNWFGCTWNNSTGFTKV	180
Db	504	GPVYCFTPSPVVVGTTRDSGAPTVSWGANDTDVFLNNTRPPLGNWFGCTWNNSTGFTKV	563
Qy	181	CGAPPCVIGGAGNNTLHCPTDCFRRKHDPATVSRGSGSPWITPRCLVDVYPRLWHYPCTIN	240
Db	564	CGAPPCVIGGVGNNTLLCPTDCFRRKHPEATVSRGSGSPWITPRCMVDVYPRLWHYPCTIN	623
Qy	241	YTIKIRMYGVGVEHRLAEACNWTGRGCDUEDDRSELSPLLLTTTQOWVLPSCFTTLP	300
Db	624	YTIKIRMYGVGVEHRLAEACNWTGRGCDUEDDRSELSPLLLSTTQOWVLPSCFTTLP	683
Qy	301	ALSTGLIHLHQNIVDVQVLYGVGSSIASWALKWEYVVLFLLLADARVCSCLLWMLLISQ	360
Db	684	ALSTGLIHLHQNIVDVQVLYGVGSSIASWALKWEYVVLFLLLADARVCSCLLWMLLISQ	743
Qy	361	AEA 363	
Db	744	AEA 746	

RESULT 4

ID	036608	PRELIMINARY;	PRT;	3011 AA.
AC	036608;			
DT	01-JAN-1998	(TREMBLrel. 05, Created)		
DT	01-JAN-1998	(TREMBLrel. 05, Last sequence update)		
DT	01-OCT-2000	(TREMBLrel. 15, Last annotation update)		
DE	POLYPROTEIN.			
OS	Hepatitis C virus strain H77.			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;			
OC	Hepacivirus.			
OX	NCBI_TaxID=63746;			
RN	[1]			
RC	SEQUENCE FROM N.A.			
RC	STRAIN=H77;			
RX	MEDLINE=97385173; PubMed=9238047;			
RA	Yanagi M., Purcell R.H., Emerson S.O., Bukh J.;			
RT	"Transcripts from a single full-length cDNA clone of hepatitis C virus			
RT	are infectious when directly transfected into the liver of a			
RT	chimpanzee.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 94:8738-8743(1997).			
DR	EMBL; AF011751; AAB67036.1; -			
DR	INTERPRO: IPR000745; -			
DR	INTERPRO: IPR001410; -			
DR	INTERPRO: IPR001490; -			
DR	INTERPRO: IPR002166; -			
DR	INTERPRO: IPR002518; -			
DR	INTERPRO: IPR002519; -			
DR	INTERPRO: IPR002521; -			
DR	INTERPRO: IPR002522; -			
DR	INTERPRO: IPR002531; -			
DR	INTERPRO: IPR002868; -			
DR	PFAM: PF00998; HCV_RdRP; 1.			
DR	PFAM: PF01001; HCV_NS4b; 1.			
DR	PFAM: PF01006; HCV_NS4a; 1.			
DR	PFAM: PF01506; HCV_NS5a; 1.			
DR	PFAM: PF01538; HCV_NS2; 1.			

DR	PFAM; PF01539; HCV_env; 1.	
DR	PFAM; PF01542; HCV_core; 1.	
DR	PFAM; PF01543; HCV_capsid; 1.	
DR	PFAM; PF01560; HCV_NSl; 1.	
DR	PRODOM; PD186062; -; 1.	
DR	polyprotein.	
SK	SEQUENCE	3011 AA; 327114 MW;
QW		0B75E6B81CB95C198 CRC64;

Query Match	94.6%	Score 1946;	DB 12;	Length 3011;
Best Local Similarity	93.7%	Pred. No. 4.6e-179;		
Matches 340;	Conservative 11;	Mismatches 12;	Indels 0;	Gaps 0;

Qy	1	ETHVTSAGHTVSGFVSLIAPGAKQNVQLINTNGSHLNSTALNCNDSLNTGWLACLFY	60
Db	384	ETHVTGNAGHTTASGLVGLITPGAKQNIQLINTNGSHIINSTALNCNESLNTGWLACLFY	443
Qy	61	HHKFNSSGCPERLASCRPLTDFQGWGCPISYANGSGDORPYCHVHPKPCGIVPAKSVC	120
Db	444	QHKFNSSGCPERLASCRRLTDFQGWGCPISYANGSGUDERPYPCHVPPRCPGIVPAKSVC	503
Qy	121	GPVYCFPTSPVVVGTTRDSGAPTSWGENDTDVFLNNTRPPLGNWFGCTWMNSTGFTKV	180
Db	504	GPVYCFPTSPVVVGTTRDSGAPTSWGANDTDVFLNNTRPPLGNWFGCTWMNSTGFTKV	563
Qy	181	CGAPPCVIGGAGNNTLHCPTDCFRKHPDATYSRCGSGPWITPRCLVDYPYRLWHYPCTIN	240
Db	564	CGAPPCVIGGVGNNTLLCPTDCFRKHPDATYSRCGSGPWITPRCMVDYPYRLWHYPCTIN	623
Qy	241	YTFIKRMVYGGVEHRLAEACNNTRGERCDLDRDRSELSPLLLTTTQOVLPCSFSTTLP	300
Db	624	YTFIKRMVYGGVEHRLAEACNNTRGERCDLDRDRSELSPLLLSTTQOVLPCSFSTTLP	683
Qy	301	ALSTGLTLHQNIVDVQYLYXGSSSTASWAIKWEYVYVLLFLLLLADARVCSCLMMMLLSQ	360
Db	684	ALSTGLTLHQNIVDVQYLYXGSSSTASWAIKWEYVYVLLFLLLLADARVCSCLMMMLLSQ	743
Qy	361	AEA 363	
Db	744	AEA 746	

RESULT

ID	Q03463	PRELIMINARY;	PRT;	3011 AA.
AC	Q03463;			
DT	01-NOV-1996	(TREMBLrel. 01, Created)		
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)		
DT	01-OCT-2000	(TREMBLrel. 15, Last annotation update)		
DE	GENOME POLYPROTEIN.			
OS	Hepatitis C virus.			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae.			
OC	Hepacivirus.			
OX	NCBI_Taxid=11103;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=93117120; PubMed=1335573;			
RA	Okamoto H., Kanai N., Mishiro S.;			
RT	Full-length nucleotide sequence of a Japanese hepatitis C virus			
RT	isolate (HC-J1) with high homology to USA isolates.;			
RL	Nucleic Acids Res. 20:6410-6410(1992).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=91013116; PubMed=2170712;			
RA	Okamoto H., Okada S., Sugiyama Y., Yotsumoto S., Tanaka T.,			
RA	Yoshizawa H., Tsuda F., Miyakawa Y., Mayumi M.;			
RT	The 5'-terminal sequence of the hepatitis C virus genome.;			
RT	Jpn. J. Exp. Med. 60:167-177(1990).			
DR	EMBL;	D10749;	BAA01582.1;	-
DR	HSSP;	P27958;	1HEI	-
DR	INTERPRO;	IPR000745;	-	-
DR	INTERPRO;	IPR001410;	-	-
DR	INTERPRO;	IPR001490;	-	-

[1]
RN SEQUENCE FROM N.A.
RP STRAIN=H77;
RX MEDLINE=97385173; PubMed=9238047;
RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;
RT "Transcripts from a single full-length cDNA clone of hepatitis C virus
RT are infectious when directly transfected into the liver of a
RT chimpanzee.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:8738-8743(1997).
DR EMBL: AF011752; AAB67037.1; -;
DR HSSP: P27938; IHEI.
DR INTERPRO: IPR000745; -;
DR INTERPRO: IPR001410; -;
DR INTERPRO: IPR001490; -;
DR INTERPRO: IPR002166; -;
DR INTERPRO: IPR002518; -;
DR INTERPRO: IPR002519; -;
DR INTERPRO: IPR002521; -;
DR INTERPRO: IPR002531; -;
DR INTERPRO: IPR002522; -;
DR INTERPRO: IPR002531; -;
DR PFAM: PF00998; HCV_RdRP; 1.
DR PFAM: PF01001; HCV_NS4b; 1.
DR PFAM: PF01006; HCV_NS4a; 1.
DR PFAM: PF01506; HCV_NS4a; 1.
DR PFAM: PF01538; HCV_NS2; 1.
DR PFAM: PF01539; HCV_NS2; 1.
DR PFAM: PF01542; HCV_core; 1.
DR PFAM: PF01543; HCV_core; 1.
DR PFAM: PF01560; HCV_capsid; 1.
DR PRODOM: PD186062; -; 1.
KW Polyprotein.
SQ SEQUENCE 3011 AA; 327262 MW; 10D1C9702CA9B5DC CRC64;

Query Match 93.9%; Score 1932; DB 12; Length 3011;
Best Local Similarity 93.4%; Pred. No. 1e-177;
Matches 339; Conservative 9; Mismatches 15; Indels 0; Gaps 0;

Qy 1 ETHVTVGSAGHTVSGFVSLAPGAKQNVOLINTNGSWHLNLTALNCNDSLTGWLGLFY 60
Db 384 ETHVTVGSAGHTVSGFVSLAPGAKQNVOLINTNGSWHLNLTALNCNDSLTGWLGLFY 443

Qy 61 HHKNSGGCERLASCRPLTDFDQGWGPISYANGSGDPDRPCYWHYPPKPGIVPAKSVC 120
Db 444 RHKNSGGCERLASCRPLTDFDQGWGPISYANGSGDLDRPCYWHYPPKPGIVPAKSVC 503

Qy 121 GPVYCFPSVVGTTDRSGAPTSYANGNDTDFVNLNTRPPLGNWFGCTWMNSTGFTKV 180
Db 504 GPVYCFPSVVGTTDRSGAPTSYANGNDTDFVNLNTRPPLGNWFGCTWMNSTGFTKV 563

Qy 181 CGAPPCVIGGAGNNTLHCPTDCFRKHPEATYSRCGSGPWITPRCLVDYPRYLWHYPCNTIN 240
Db 564 RVGVCVIGGVGNLTLLCPTDCFRKHPEATYSRCGSGPWITPRCLVDYPRYLWHYPCNTIN 623

Qy 241 YTFIKRMVYGVGVEHLEAACNWTGERCDLDRDSELSPLLLTTQVQLPCSFITLP 300
Db 624 YTFIKRMVYGVGVEHLEAACNWTGERCDLDRDSELSPLLLTTQVQLPCSFITLP 683

Qy 301 ALSTGLIHLHQNVDVQVLYGVGSSIASWAIKWEYVLLFLLADARVCSCLMWLLISQ 360
Db 684 ALSTGLIHLHQNVDVQVLYGVGSSIASWAIKWEYVLLFLLADARVCSCLMWLLISQ 743

Qy 361 AEA 363
Db 744 AEA 746

RESULT 8
Q04185
ID Q04185 PRELIMINARY; PRT: 778 AA.
AC Q04185: Q01811;
DT 01-NOV-1998 (Tremblrel. 08, Created)

DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C
DE (CORE PROTEIN); MATRIX PROTEIN
DE (ENVELOPE PROTEIN M); MAJOR ENVELOPE PROTEIN E; NONSTRUCTURAL PROTEIN
DE NS1: NONSTRUCTURAL PROTEIN NS2; NONSTRUCTURAL PROTEIN NS4B;
DE NS1: NONSTRUCTURAL PROTEIN NS4B; HELICASE
DE (NS3); RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (NS5)] (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91195357; PubMed=1849654;
RA Ogata N., Alter H.J., Miller R.H., Purcell R.H.;
RT "Nucleotide sequence and mutation rate of the H strain of hepatitis C
RT virus.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:3392-3396(1991).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION. NS3
CC AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
DR EMBL: M62382; AAB02128.1; -;
DR INTERPRO: IPR002518; -;
DR INTERPRO: IPR002519; -;
DR INTERPRO: IPR002521; -;
DR INTERPRO: IPR002522; -;
DR INTERPRO: IPR002531; -;
DR PFAM: PF01538; HCV_NS2; 1.
DR PFAM: PF01539; HCV_env; 1.
DR PFAM: PF01542; HCV_core; 1.
DR PFAM: PF01543; HCV_capsid; 1.
DR PFAM: PF01560; HCV_NS1; 1.
KW Polyprotein; Glycoprotein; RNA-directed RNA polymerase; Core protein;
KW Coat protein; Envelope protein; Helicase; ATP-binding; Transmembrane;
KW Nonstructural protein.
FT INIT_MET 1
FT CHAIN 2 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 729
FT CHAIN 730 >778
FT TRANSMEM 347 369
FT NON_TER 778 778
FT SEQUENCE 778 AA; 85530 MW; 6E37E50AB97B6C32 CRC64;

Query Match 93.5%; Score 1925; DB 12; Length 778;
Best Local Similarity 91.7%; Pred. No. 1e-177;
Matches 333; Conservative 13; Mismatches 17; Indels 0; Gaps 0;

Qy 1 ETHVTVGSAGHTVSGFVSLAPGAKQNVOLINTNGSWHLNLTALNCNDSLTGWLGLFY 60
Db 384 ETHVTVGSAGHTVSGFVSLAPGAKQNVOLINTNGSWHLNLTALNCNDSLTGWLGLFY 443

Qy 61 HHKNSGGCERLASCRPLTDFDQGWGPISYANGSGDPDRPCYWHYPPKPGIVPAKSVC 120
Db 444 HHKNSGGCERLASCRPLTDFDQGWGPISYANGSGDPDRPCYWHYPPKPGIVPAKSVC 503

Qy 121 GPVYCFPSVVGTTDRSGAPTSYANGNDTDFVNLNTRPPLGNWFGCTWMNSTGFTKV 180
Db 504 GPVYCFPSVVGTTDRSGAPTSYANGNDTDFVNLNTRPPLGNWFGCTWMNSTGFTKV 563

Qy 181 CGAPPCVIGGAGNNTLHCPTDCFRKHPEATYSRCGSGPWITPRCLVDYPRYLWHYPCNTIN 240
Db 564 CGAPPCVIGGAGNNTLHCPTDCFRKHPEATYSRCGSGPWITPRCLVDYPRYLWHYPCNTIN 623

QY 241 YTIKIRMYGGVEHRLAACAACNWTGRGCDLEDRDRSELSPLLLTTTQWQVLPSCFTTLP 300
DB 624 YTIKIRMYGGVEHRLAACAACNWTGRGCDLEDRDRSELSPLLLTTTQWQVLPSCFTTLP 683
QY 301 ALSTGLIHLHQNIVDVQVLYGVGSSIASWAIKWEYVVLLELLADARVCSCLMWMLLSQ 360
DB 684 ALSTGLIHLHQNIVDVQVLYGVGSSIASWAIKWEYVVLLELLADARVCSCLMWMLLSQ 743
QY 361 AEA 363
DB 744 AEA 746
RESULT 9
ID Q9PXZ2 PRELIMINARY; PRT: 403 AA.
AC Q9PXZ2;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
DE E2/NS1 REGION.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92348860; Pubmed=1668329;
RA Kreamsdorf D., Porchon C., Brechot C.;
RT "Hepatitis C virus (HCV)-RNA in non-A, non-B chronic hepatitis in
J. France. Nucleotide sequence of a French HCV isolate.";
RL J. Hepatol. 13:24-24(1991).
DR INTERPRO; IPR002519;
DR INTERPRO; IPR002531;
DR PFAM; PF01539; HCV_env; 1.
DR PFAM; PF01560; HCV_NSI; 1.
SQ SEQUENCE 403 AA; 44565 MW; 06278192EAC5B3F1 CRC64;

Query Match 88.3%; Score 1817; DB 12; Length 403;
Best Local Similarity 86.7%; Pred. No. 1.3e-167;
Matches 306; Conservative 30; Mismatches 17; Indels 0; Gaps 0;

QY 1 ETHVTGCSAGHTVSGFVSLAPCAKONVOLINTNGSWHLNSTALNCNDSLTGWLGLFY 60
DB 51 EYITGGSTARTQGLVSLFSGAKQDILQINTNGSWHLNSTALNCNDSLTGWLGLFY 110
QY 61 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDPQRPYCHWHPKPGIYPAKSV 120
DB 111 YHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDPQRPYCHWHPKPGIYPAKSV 170
QY 121 GPVYCTPSPVVVGTDRSGAPTSYANGSGDPQRPYCHWHPKPGIYPAKSV 180
DB 171 GPVYCTPSPVVVGTDRSGAPTSYANGSGDPQRPYCHWHPKPGIYPAKSV 230
QY 181 CGAPPCVIGGAGNNTLHCPTDCFRKHDPATYSRCGSGPWTIPRCLVDPYRLWHYPCTIN 240
DB 231 CGAPPCVIGGAGNNTLYCPTDCFRKHDPATYSRCGSGPWTIPRCLVDPYRLWHYPCTIN 290
QY 241 YTIKIRMYGGVEHRLAACAACNWTGRGCDLEDRDRSELSPLLLTTTQWQVLPSCFTTLP 300
DB 291 YTIKIRMYGGVEHRLAACAACNWTGRGCDLEDRDRSELSPLLLTTTQWQVLPSCFTTLP 350
QY 301 ALSTGLIHLHQNIVDVQVLYGVGSSIASWAIKWEYVVLLELLADARVCSCLMWMLLSQ 353
DB 351 ALSTGLIHLHQNIVDVQVLYGVGSSIASWAIKWEYVVLLELLADARVCSCLMWMLLSQ 403
RESULT 10
ID Q9J3H0 PRELIMINARY; PRT: 3010 AA.
AC Q9J3H0;
DT 01-OCT-2000 (Tremblrel. 15, Created)

DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE POLYPROTEIN.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MD22;
RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT "Characteristics of hepatitis C viral genome associated with disease
progression.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF207763; AAF65953.1;
SQ SEQUENCE 3010 AA; 327098 MW; 737EEF31E3C2B28D CRC64;

Query Match 87.2%; Score 1794; DB 12; Length 3010;
Best Local Similarity 83.5%; Pred. No. 2.3e-164;
Matches 303; Conservative 30; Mismatches 30; Indels 0; Gaps 0;

QY 1 ETHVTGCSAGHTVSGFVSLAPCAKONVOLINTNGSWHLNSTALNCNDSLTGWLGLFY 60
DB 384 DTHVTGGVYAHYTTQGFASLFAFGASQNIQLINTNGSWHLNSTALNCNDSLTGWLGLFY 443
QY 61 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDPQRPYCHWHPKPGIYPAKSV 120
DB 444 THEFNASSGCPERLASCRPIQDFDQGWGPIYTPNSPDQRPYCHWHPKPGIYPAKSV 503
QY 121 GPVYCTPSPVVVGTDRSGAPTSYANGSGDPQRPYCHWHPKPGIYPAKSV 180
DB 504 GPVYCTPSPVVVGTDRSGAPTSYANGSGDPQRPYCHWHPKPGIYPAKSV 563
QY 181 CGAPPCVIGGAGNNTLHCPTDCFRKHDPATYSRCGSGPWTIPRCLVDPYRLWHYPCTIN 240
DB 564 CGAPPCVIGGAGNNTLHCPTDCFRKHDPATYSRCGSGPWTIPRCLVDPYRLWHYPCTIN 623
QY 241 YTIKIRMYGGVEHRLAACAACNWTGRGCDLEDRDRSELSPLLLTTTQWQVLPSCFTTLP 300
DB 624 YTIKIRMYGGVEHRLAACAACNWTGRGCDLEDRDRSELSPLLLTTTQWQVLPSCFTTLP 683
QY 301 ALSTGLIHLHQNIVDVQVLYGVGSSIASWAIKWEYVVLLELLADARVCSCLMWMLLSQ 360
DB 684 ALSTGLIHLHQNIVDVQVLYGVGSSIASWAIKWEYVVLLELLADARVCSCLMWMLLSQ 743
QY 361 AEA 363
DB 744 AEA 746

RESULT 11
ID Q81754 PRELIMINARY; PRT: 3011 AA.
AC Q81754;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE POLYPROTEIN.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-G9;
RA Okamoto H.;
RL Submitted (APR-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-G9;
RX MEDLINE=94172337; Pubmed=8126459;
RA Okamoto H., Kojima M., Sakamoto M., Iizuka H., Hagiwara S.,

[illegible][illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 11:51:45 ; Search time 47.68 Seconds
(without alignments)
182.157 Million cell updates/sec

Title: US-09-407-430-3
Perfect score: 1494
Sequence: 1 GAKONVQLNTGSHLNT.....LEACNWTGRCDELDNR 254

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues
Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_36.*
1: /SIDS1/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq/AA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1494	100.0	402	14 R34438	Sequence of glycoprotein
2	1494	100.0	480	14 R33992	HCV-1 E2/NSI prote
3	1494	100.0	663	17 R92935	HCV1 E2 + NS2 poly
4	1494	100.0	663	20 W67615	Hepatitis C virus
5	1494	100.0	2772	11 R08123	Hepatitis C virus
6	1494	100.0	2955	20 Y14975	Amino acid sequenc
7	1494	100.0	3011	13 R21519	Compiled HCV seque
8	1494	100.0	3011	17 R90931	Hepatitis C virus
9	1494	100.0	3011	18 W34480	HCV polyprotein.
10	1494	100.0	3011	19 W40038	HCV polyprotein.
11	1492	99.9	3011	14 R31621	Hepatitis C virus
12	1489	99.7	2955	11 R08124	Hepatitis C virus

13	1486	99.5	2816	14 R34009	HCV-1 polyprotein.
14	1473	98.6	621	14 R33185	Sequence of subfra
15	1473	98.6	622	14 R33591	HCV CKS-NS1S1-NS1S
16	1473	98.6	738	14 R33592	HCV CKS-full lengt
17	1473	98.6	2894	13 R24440	Composite HCV HC-J
18	1473	98.6	2894	16 R70230	Composite hepatitis
19	1438	96.3	3011	15 R66995	Hepatitis C virus
20	1426	95.4	402	14 R34439	Sequence of glycop
21	1426	95.4	409	14 R33995	H77 E2/NSI protein
22	1426	95.4	733	14 R38278	NANB hepatitis vir
23	1419	95.0	1006	18 W12715	HCV genome type 1a
24	1418	94.9	3011	19 W77397	Hepatitis C virus
25	1418	94.9	3011	20 W98020	Infectious hepatitis
26	1416	94.8	3011	19 W77398	Hepatitis C virus
27	1414	94.6	402	14 R34440	Sequence of glycop
28	1414	94.6	409	14 R33996	H90 E2/NSI protein
29	1413	94.6	305	18 W00929	Recombinant HCV E2
30	1413	94.6	333	14 R40118	HGH-HCV-E2 fusion
31	1413	94.6	367	16 R79218	pHCV167-encoded pr
32	1413	94.6	359	14 R40117	HGH-HCV-E2 fusion
33	1413	94.6	537	14 R40114	APP-HCV-E2 fusion
34	1413	94.6	967	16 R79222	pHCV141-encoded se
35	1413	94.6	1648	16 R79221	pHCV176-encoded se
36	1413	94.6	3011	14 R40119	HCV genomic amino
37	1413	94.6	3011	16 R79232	HCV sequence. Hep
38	1407	94.2	367	14 R40115	APP-HCV-E2 fusion
39	1397	93.5	3011	14 R40120	HCV genomic amino
40	1390	93.0	3011	13 R22154	NANBV Huch c59 is
41	1389	93.0	278	14 R33997	Th E2/NSI protein.
42	1386	92.8	337	16 R79217	pHCV351-encoded AP
43	1386	92.8	377	16 R79226	pHCV422-encoded AP
44	1386	92.8	397	16 R79220	pHCV419-encoded AP
45	1386	92.8	410	16 R79227	pHCV423-encoded AP

ALIGNMENTS

RESULT 1	
R34438	
ID R34438 standard; Protein; 402 AA.	
XX	
AC R34438;	
XX	
DT 09-AUG-1993 (first entry)	
XX	
DE Sequence of glycoprotein E2/NSI in clone HCV1.	
XX	
KW Hepatitis C virus; envelope protein; glycoprotein; E2/NSI;	
KW diagnostic reagent.	
XX	
OS Hepatitis C virus.	
XX	
PN EP537626-A.	
XX	
PD 21-APR-1993.	
XX	
PF 08-OCT-1992; 92EP-0117191.	
XX	
PR 08-OCT-1991; 91JP-0260824.	
XX	
PA (NAHE-) NAT INST OF HEALTH.	
XX	
PI Harada S, Honda Y, Miyamura T, Saito I;	
XX	
DR WPI; 1993-127516/16.	
DR N-PSDB; Q40330.	
XX	
PT Diagnostic reagent for hepatitis C virus - comprises second	
PT envelope protein or first non-structural protein encoded by HCV	
XX	
PS gene and has sugar chain	
XX	
PS Claim 2; Pages 30-32; 58pp; English.	

```

XX CC Glycoprotein E2/NS1 is derived from the second envelope protein or
XX CC first non-structural protein encoded by the genome of HCV. The
XX CC nucleic acid is extracted from the serum of the patient of hepatitis
XX CC C. The serum is pref. mixed with transfer RNA (tRNA) as a carrier
XX CC of virus RNA. As a technique of cloning cDNA from the nucleic acid,
XX CC it is preferred to use polymerase chain reaction method. In the
XX CC reaction, any commercially available random primers or synthesized
XX CC DNA having a base sequence similar to that of primer AS1 may be used
XX CC as a primer. Representative examples of sense primers include S1.
XX SQ
SQ Sequence 402 AA;

Query Match 100.0%; Score 1494; DB 14; Length 402;
Best Local Similarity 100.0%; Pred. No. 7.8e-125; Indels 0; Gaps 0;
Matches 254; Conservative 0; Mismatches 0;

QY 1 GAKONVOLINTNGSWHLNSTALNCNDSLNTGWLGLFYYHHKFNSSGCPERLASCRPLTDF 60
DB 67 gakqvqlntngswhlntstalnclndslntgwlglfyyhhkfnssgcperscrpltdf 126

QY 61 DQGWGPTISYANGSGDQRPYCWHPKPGCGIVPAKSVCGPVYCFTPSPVVVGTTRSGAP 120
DB 127 dqgwgptisyangsgdqrpycwhypkpgcgvlpaksvcgpvycftpsvrvvgtttrsgap 186

QY 121 TYSWGENDTDVFLNTRPPLGNWFGCTWNNSTGFTKVCGAPPCVIGGAGNNTLHCPTDC 180
DB 187 tyswendtdvflntrpplgnwfgctwmnstgftkvcgappcvigaggnntlhcptdc 246

QY 181 FRKHPDATYRCGSGPWITPRCLVDYPRYLWHYPCTINYTIKIRMYVGGVEHRLAECN 240
DB 247 frkhpdatysrcsgsgpwitprclvdyprrlywhypctinytifkirmvvggvehrleaecn 306

QY 241 WTRGERCDLEDRDR 254
DB 307 wtrgercdledrdr 320

RESULT 2
R33992 ID R33992 standard; Protein; 480 AA.
XX AC
XX AC R33992;
XX DT 26-JUL-1993 (first entry)
XX DE
XX DE HCV-1 E2/NS1 protein.
XX KW Polymerase chain reaction; PCR; amplify; primer; hepatitis C virus;
XX KW HCV; asymptomatic; chronically infected; epitope; viral isolate;
XX KW domain; immunological; cross-reactive; envelope protein; vaccine;
XX KW gp53(BVDV)/gp55; hog cholera virus; pestivirus; NS1; flavivirus.
XX OS
XX OS Synthetic.
XX PN WO9306126-A.
XX XX
XX PD 01-APR-1993.
XX PF 11-SEP-1992; 92WO-US07683.
XX PR 13-SEP-1991; 91US-0759575.
XX XX (CHIR ) CHIRON CORP.
XX XX Houghton M, Weiner AJ;
XX XX WPI; 1993-117468/14.
XX DR
XX PT Immuno-reactive hepatitis C virus polypeptide compsns. - contg.
XX PT at least 2 sequences from the first variable domain of distinct
XX PT HCV isolates

```

```

XX PS Disclosure; Fig 3; 106pp; English.
XX CC
XX CC The sequences given in R33992-002 represent a portion of the E2/NS1
XX CC protein encoded by group I and group II HCV isolates, from amino acid
XX CC 370-850. E2/NS1 are viral envelope proteins and are of immunogenic
XX CC interest. E2/NS1 contains an N-terminal hypervariable domain of about
XX CC 30 amino acids which shows large variation between nearly all isolates.
XX CC This is an important immunoreactive domain. This putative envelope
XX CC glycoprotein E2/NS1 may correspond to the gp53(BVDV)/gp55 (hog cholera
XX CC virus) envelope polypeptide of the pestiviruses and the NS1 of the
XX CC flaviviruses, both of which confer protective immunity in hosts
XX CC vaccinated with these polypeptides. It has been discovered that a
XX CC number of important HCV epitopes vary among viral isolates and that
XX CC these epitopes can be mapped to specific domains. This meant that
XX CC immunologically cross-reactive polypeptides which focus on variable
XX CC rather than constant domains can be produced. See also Q39134-48
XX CC and R33982-91.
XX SQ
SQ Sequence 480 AA;

Query Match 100.0%; Score 1494; DB 14; Length 480;
Best Local Similarity 100.0%; Pred. No. 9.6e-125; Indels 0; Gaps 0;
Matches 254; Conservative 0; Mismatches 0;

QY 1 GAKONVOLINTNGSWHLNSTALNCNDSLNTGWLGLFYYHHKFNSSGCPERLASCRPLTDF 60
DB 37 gakqvqlntngswhlntstalnclndslntgwlglfyyhhkfnssgcperscrpltdf 96

QY 61 DQGWGPTISYANGSGDQRPYCWHPKPGCGIVPAKSVCGPVYCFTPSPVVVGTTRSGAP 120
DB 97 dqgwgptisyangsgdqrpycwhypkpgcgvlpaksvcgpvycftpsvrvvgtttrsgap 156

QY 121 TYSWGENDTDVFLNTRPPLGNWFGCTWNNSTGFTKVCGAPPCVIGGAGNNTLHCPTDC 180
DB 157 tyswendtdvflntrpplgnwfgctwmnstgftkvcgappcvigaggnntlhcptdc 216

QY 181 FRKHPDATYRCGSGPWITPRCLVDYPRYLWHYPCTINYTIKIRMYVGGVEHRLAECN 240
DB 217 frkhpdatysrcsgsgpwitprclvdyprrlywhypctinytifkirmvvggvehrleaecn 276

QY 241 WTRGERCDLEDRDR 254
DB 277 wtrgercdledrdr 290

RESULT 3
R33935 ID R33935 standard; Protein; 663 AA.
XX AC
XX AC R33935;
XX DT 04-JUN-1996 (first entry)
XX DE
XX DE HCV1 E2 + NS2 polypeptide.
XX KW HCV1; E1 protein; E2 protein; truncation; vaccine; diagnosis;
XX KW therapy; protein secretion.
XX OS
XX OS Hepatitis c virus.
XX XX
XX XX Key Location/Qualifiers
XX FT Peptide 1..21 /label= Sig_peptide
XX FT Protein 22..336 /label= Mat_protein
XX FT /note= "E2 mature protein"
XX FT Domain 337..361 /label= Membrane-spanning_domain
XX FT /note= "E2 C-terminal anchor"
XX FT Protein 362..663 /label= Mat_protein

```



```

FT XX /note= "NS2A+NS2B proteins"
PN XX W09604301-A2.
XX 15-FEB-1996.
XX 26-JUL-1995; 95WO-US10035.
XX 25-JUL-1995; 95US-0282959.
PR 29-JUL-1994; 94US-0282959.
XX (CHIR ) CHIRON CORP.
XX Houghton M, Selby M;
PI WPI: 1996-129331/13.
DR N-PSDB; T18246.
XX Secretable hepatitis C virus E1 and E2 polypeptide(s) lacking
PT all/part of the membrane spanning domain - useful in vaccines, and
PT for diagnostic and therapeutic purposes, e.g. in assays for HCV
XX PS Disclosure: Fig 4a-4c; 46pp; English.
XX Truncated versions of the hepatitis C virus E2 envelope protein
CC (R92935) are obtd. by deletion the membrane-spanning C-terminal
CC anchor domain (approx. amino acids 337-361) of the protein.
CC This can be achieved by PCR amplification of the E2 gene (T18246)
CC using appropriate primers. Truncated versions of the E1
CC polypeptide (see R92934) are similarly produced. When produced
CC recombinantly in host cells, the truncated proteins are secreted
CC into the medium. When co-expressed or combined after separate
CC expression, the truncated E2 and E1 proteins form a complex. The
CC truncated proteins and complex are useful in vaccines and for
CC diagnostic and therapeutic appls.
XX SQ Sequence 663 AA:
Query Match 100.0%; Score 1494; DB 17; Length 663;
Best Local Similarity 100.0%; Pred. No. 1.4e-124;
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAKQNVQLINTNGSHLNTALNCNDLSNTGLAGLFYHHKFNSSGCPERLASCRPLTDF 60
DB 43 gaknvqlintngshlnstalnclndslntglaglfyhhkfnssgcperscrpltdf 102
QY 61 DQGWGPISYANGSGDPQRPYCHWHYPKPCGIVPAKSVCGPVYCFPTSPVVVGTDRSGAP 120
DB 103 dggwgpisyangsgdpqrpypchwhypkpcgipkavscgvpvycftspvvvgttdrsgap 162
QY 121 TYSWGENDTDFVFLNTRPPLGNWFGCTWMNSTGFTKVCAGAPPCVYGGAGNNTLHCPTDC 180
DB 163 tyswgendtdvflntrpplgnwfgctwmnstgftkvcgappcvlggagntnlhcptdc 222
QY 181 FRKHDPATYSRCGSGPWITPRCLVDYPYRLWHYPCPTINTYFIKIRMYVGVGVEHRLFAACN 240
DB 223 frkhdpatysrcsgspwiltprclvdypyrllwhypcctintyfiikirmvvgvghrleaaacn 282
QY 241 WTRGERCDLEDRDR 254
DB 283 wtrgercdledrdr 296
RESULT 4
W67615
ID W67615 standard; Protein; 663 AA.
XX W67615;
XX 02-MAR-1999 (first entry)
XX Hepatitis C virus E2 protein.

```

```

XX Hepatitis C virus; E2; HCV; truncation; variant; PCR; amplification;
KW affinity chromatography; Galanthus nivalis; agglutinin; medicament;
XX detection; infection.
XX Hepatitis C virus.
OS W09850556-A2.
PN 12-NOV-1998.
XX 06-MAY-1998; 98WO-US09097.
XX 06-MAY-1997; 97US-0045675.
PR (CHIR ) CHIRON CORP.
XX PA
XX ABrignani S, Chien D, Choo QL, Glazer E, Houghton M;
PI Selby M;
XX WPI: 1999-034724/03.
DR N-PSDB; V81371.
XX Methods for isolating truncated HCV E1 and E2 polypeptides - used
PT in, e.g. immunodiagnostic kits for diagnosis of HCV infection
XX PS Disclosure: Fig 2A-E; 65pp; English.
XX This sequence represents the Hepatitis C virus (HCV) E2 protein. The
CC invention relates to the intracellular production and isolation of
CC C-terminally truncated variants of the E2 protein. The truncations
CC start from about amino acid residue 500 of the E2 protein and are
CC generated by PCR amplification of the gene sequence. The truncated
CC proteins are then produced intracellularly in host cells which are
CC disrupted. The HCV proteins are then purified preferably by affinity
CC chromatography, especially using Galanthus nivalis agglutinin resin.
CC The HCV E1 and HCV E2 polypeptides can be used to manufacture a
CC medicament useful for detecting the presence or absence of HCV infection
CC in an individual. They can also be used in a immunodiagnostic test kit
CC for detecting HCV infection.
XX SQ Sequence 663 AA:
Query Match 100.0%; Score 1494; DB 20; Length 663;
Best Local Similarity 100.0%; Pred. No. 1.4e-124;
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAKQNVQLINTNGSHLNTALNCNDLSNTGLAGLFYHHKFNSSGCPERLASCRPLTDF 60
DB 43 gaknvqlintngshlnstalnclndslntglaglfyhhkfnssgcperscrpltdf 102
QY 61 DQGWGPISYANGSGDPQRPYCHWHYPKPCGIVPAKSVCGPVYCFPTSPVVVGTDRSGAP 120
DB 103 dggwgpisyangsgdpqrpypchwhypkpcgipkavscgvpvycftspvvvgttdrsgap 162
QY 121 TYSWGENDTDFVFLNTRPPLGNWFGCTWMNSTGFTKVCAGAPPCVYGGAGNNTLHCPTDC 180
DB 163 tyswgendtdvflntrpplgnwfgctwmnstgftkvcgappcvlggagntnlhcptdc 222
QY 181 FRKHDPATYSRCGSGPWITPRCLVDYPYRLWHYPCPTINTYFIKIRMYVGVGVEHRLFAACN 240
DB 223 frkhdpatysrcsgspwiltprclvdypyrllwhypcctintyfiikirmvvgvghrleaaacn 282
QY 241 WTRGERCDLEDRDR 254
DB 283 wtrgercdledrdr 296
RESULT 5
R08123
ID R08123 standard; protein; 2772 AA.
XX

```

AC R08123;
XX 23-JAN-1991 (first entry)
XX Hepatitis C virus polypeptide from long ORF.
DE Hepatitis C virus; antiviral agent.
XX Hepatitis C virus.
XX EP388232-A.
XX 19-SEP-1990.
XX 16-MAR-1990; 90EP-0302866.
XX 18-MAY-1989; 89US-0355002.
PR 17-MAR-1989; 89US-0325338.
PR 20-APR-1989; 89US-0341334.
XX (CHIR-) CHIRON CORP.
XX Houghton M, Choo QL, Kuo G;
XX WPI; 1990-284418/38.
DR N-PSDB; Q05955.
XX Hepatitis C virus DNA - used for producing probes,
PT polypeptide(s), antibodies and anti-sense polynucleotide(s) for
PT diagnosis and therapy.
XX Disclosure; Fig 16; 83pp; English.
XX HCV cDNA libraries were constructed using pooled serum from a
CC chimpanzee with chronic HCV infection. A lambda gtl library was
CC screened with probes derived from previously isolated clones. The
CC ORF is derived from the overlapping clones p14a, CA167b, CA156e,
CC CA84a, CA59a, K9-1, 12f, 14i, 11b, 7f, 8h, 33c, 40b, 37b, 35, 36,
CC 81, 32, 33b, 25c, 14c, 8f, 33f, 33g, 39c, 35f, 19g, 26g and 15e.
CC This polypeptide can be used to design probes
CC for the detection of HCV nucleic acids, in screening programmes
CC for antiviral agents and in preparing blood free of HCV. Antisense
CC polynucleotides can be used to inhibit viral replication.
CC See also Q05956.
XX Sequence 2772 AA;
SQ
Query Match 100.0%; Score 1494; DB 11; Length 2772;
Best Local Similarity 100.0%; Pred. No. 7.6e-124;
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GAKQNVQLINTNGSWHLNSTALNCNDSINTGWLGLAGLFYHHKFNSSGCPERLASCRPLTDF 60
Db 292 gakqnvqlintngswhlnstalncnclsintgwlglaglfyhhkfnssgcperscrpltdf 351
Qy 61 DOGWPISYANGSGDQRPYCHWYPPKPGCIVPAKSVCGPVYCTPSPVVGTTDRSGAP 120
Db 352 dggwgpisyangsgdqrpychwyppkpgcivpaksvcgpyctpspvvgttdrsgap 411
Qy 121 TYSWGENDTDFVNLNTRPPLGNFGCTWMNSTGFTKVCGAPPCVIGGAGNNTLHCPTDC 180
Db 412 tyswgendtdfvlnntrpplgngfctwmnstgtfkcvcgappcvcigagantlhcptdc 471
Qy 181 FRKHPDATYSCGSGPWLTTPCLVDYDYPVRLWHYDCTINTYTFKIRMYGVGVEHLEACN 240
Db 472 frkhpdatyscgsgpwlttprclvdydypvrlwhypctintytfkirmvgygvehleacn 531
Qy 241 WTRGERCDLEDRDR 254
Db 532 wtrgercdleddr 545

RESULT 6
Y14975
ID Y14975 standard; Protein; 2955 AA.
XX
AC Y14975;
XX 08-NOV-1999 (first entry)
XX Amino acid sequence of HCV-1 ORF.
XX Hepatitis C virus; HCV; J1; J7; HCV-1; non-A, non-B HCV; NANBH;
KW HCV infection; vaccine.
XX Hepatitis C virus.
XX Key Location/Qualifiers
FT Misc-difference 441 /note= "encoded by Tt"
FT Misc-difference 461 /note= "encoded by CCCC"
XX EP939128-A2.
XX 01-SEP-1999.
XX 17-SEP-1990; 90EP-0310149.
XX 21-DEC-1989; 89US-0456142.
PR 15-SEP-1989; 89US-0408045.
XX (CHIR) CHIRON CORP.
PA (OYAA/) OYA A.
XX Cha T, Han J, Houghton M, Irvine BD, Kolberg JA;
PI Miyamura T, Saito I, Weiner AJ;
XX WPI; 1999-480843/41.
DR N-PSDB; 207656.
XX New Hepatitis C Virus isolates, useful for diagnosis of hepatitis
PT infections and development of vaccines
XX Disclosure; Fig 12; 132pp; English.
XX The invention provides two new isolates of hepatitis C virus (HCV), J1
CC and J7. These two isolates comprise nucleotide and amino acid sequences
CC that are distinct from the HCV isolate HCV-1. The nucleotide sequences
CC may be used to detect non-A, non-B HCV (NANBH) polynucleotides by
CC hybridisation for diagnosis of NANBH infections. They may also be used to
CC screen blood donors, donated blood and blood products for this infection.
CC The isolates may also be used to isolate other naturally occurring
CC variants of the virus. The polypeptides may be used as a vaccine for
CC administration to patients to protect against infection with NANBH. The
CC present sequence represents the amino acid sequence of HCV-1 ORF.
XX Sequence 2955 AA;
SQ
Query Match 100.0%; Score 1494; DB 20; Length 2955;
Best Local Similarity 100.0%; Pred. No. 8.2e-124;
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GAKQNVQLINTNGSWHLNSTALNCNDSINTGWLGLAGLFYHHKFNSSGCPERLASCRPLTDF 60
Db 406 gakqnvqlintngswhlnstalncnclsintgwlglaglfyhhkfnssgcperscrpltdf 465
Qy 61 DOGWPISYANGSGDQRPYCHWYPPKPGCIVPAKSVCGPVYCTPSPVVGTTDRSGAP 120
Db 466 dggwgpisyangsgdqrpychwyppkpgcivpaksvcgpyctpspvvgttdrsgap 525
Qy 121 TYSWGENDTDFVNLNTRPPLGNFGCTWMNSTGFTKVCGAPPCVIGGAGNNTLHCPTDC 180
Db 526 tyswgendtdfvlnntrpplgngfctwmnstgtfkcvcgappcvcigagantlhcptdc 585

QY 181 FRKHPDATYSGSGPWITPRCLVDYPRRLWHYPCTINTYIFKIRMYGVGVEHRLAECN 240
 Db 586 frkhpdatysrcsgsgpwitprclvdyprrlwhypctintyifkirmvvgvghrleaecn 645
 QY 241 WTRGERCDLEDNR 254
 Db 646 wtrgercdledrdr 659

RESULT 7
 R21519
 ID R21519 standard; Protein; 3011 AA.

XX AC R21519;

XX DT 22-JUN-1992 (first entry)

XX DE Compiled HCV sequence.

XX KW HCV1; serum; gt11.

XX OS Hepatitis C virus 1.

XX FH Key Location/Qualifiers

FT MISC-difference 9 /label= ARG
 FT MISC-difference 11 /label= THR
 FT MISC-difference 176 /label= THR
 FT MISC-difference 334 /label= VAL
 FT MISC-difference 603 /label= ILE
 FT MISC-difference 848 /label= (ASN)
 FT MISC-difference 1114 /label= SER
 FT MISC-difference 1117 /label= THR
 FT MISC-difference 1276 /label= LEU
 FT MISC-difference 1328 /label= (VAL)
 FT MISC-difference 1454 /label= TYR
 FT MISC-difference 1471 /label= (SER)
 FT MISC-difference 1877 /label= (GLY)
 FT MISC-difference 1948 /label= (HIS)
 FT MISC-difference 1949 /label= (CYS)
 FT MISC-difference 2021 /label= (VAL)
 FT MISC-difference 2349 /label= (SER)
 FT MISC-difference 2385 /label= (PHE)
 FT MISC-difference 2386 /label= (ALA)
 FT MISC-difference 2502 /label= (PHE)
 FT MISC-difference 2690 /label= (GLY)
 FT MISC-difference 2996 /label= (PRO)
 XX WO20202642-A.
 XX 20-FEB-1992.

XX PF 12-AUG-1991; 9LWO-US05728.
 XX PR 10-AUG-1990; 90US-0566209.
 XX PA (CHIR-) CHIRON CORP.
 XX PI Houghton M, Choo QL, Kuo G, Weiner AJ, Urdea MS, Irvine BD;
 PI Kolberg JA;
 XX DR WPI; 1992-080094/10.
 XX DR N-PSDB; Q21744.
 XX PT Reagents for isolating, amplifying and detecting HCV
 PT polynucleotide(s) - used to monitor spread of blood-borne non-a,
 PT non-b hepatitis virus infection and screen blood samples for
 PT virus
 XX PS Disclosure; Fig 1; 67pp; English.
 XX CC Heterogeneities in cloned DNAs of HCV1 are indicated by the amino
 CC acid indicated in the features, the parentheses indicated that the
 CC heterogeneity was detected at or near to the 5' - or 3'-end of the
 CC HCV in the clone.
 CC The sequence is derived from a composite HCV cDNA from HCV1, a
 CC prototypic HCV. The DNA sequence is based upon sequence information
 CC derived from a no. of HCV cDNA clones, which were isolated from a no.
 CC of HCV cDNA libraries, including the "c" library present in lambda
 CC gt11 (ATCC No.40394), and from human serum. The HCV cDNA clones
 CC were isolated by methods described in WO9014436.
 CC The clones from which the sequence was derived are 5'clone32,
 CC b114a, 18g, ag30a, CA205a, CA250a, CA216a, p14a, CA167b, CA156e,
 CC CA84a, CA59a, K9-1 (also called K9-1), 26j, 131, 12f, 14i, 11b, 7f,
 CC 7e, 8h, 33c, 40b, 37b, 35, 36, 81, 32, 33b, 25c, 14c, 8f, 33f, 33g,
 CC 39c, 35f, 19g, 26g, 15e, b5a, 16jh, 6k and p131jh.
 XX SQ Sequence 3011 AA;

Query Match 100.0%; Score 1494; DB 13; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 8.3e-124;
 Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAKQNVQLINTNGSWHLNSTALNCNDSLNTGWLGLFYHHKFNSSCCPERLASCRPLTDF 60
 Db 406 gakqnvqlintngswhlNSTALNCNDSLNTGWLGLFYHHKFNSSCCPERLASCRPLTDF 465
 QY 61 DOGWGPISYANGSGDPORPCYWHYPPKCGIVPAKSGVPYCFPTSPVVVGTDRSGAP 120
 Db 466 dggwgpisyangsgdpqrpycwhypkpcgylpaksvcgpycftspvvvgtdrsgap 525
 QY 121 TYSWGENDTDFVLNNTRPLGNWFGCTWMNSTGFTKVCAGPPCVIGGAGNNTLHCPTDC 180
 Db 526 tyswgenndtfdvlnntrpplgnwfgctwmnstgftkvcagppcvlggagntllhcptdc 585
 QY 181 FRKHPDATYSGSGPWITPRCLVDYPRRLWHYPCTINTYIFKIRMYGVGVEHRLAECN 240
 Db 586 frkhpdatysrcsgsgpwitprclvdyprrlwhypctintyifkirmvvgvghrleaecn 645
 QY 241 WTRGERCDLEDNR 254
 Db 646 wtrgercdledrdr 659

RESULT 8
 R90931
 ID R90931 standard; Protein; 3011 AA.
 XX AC R90931;
 XX DT 15-MAY-1996 (first entry)
 XX DE Hepatitis C virus polyprotein.

```

XX Non-A non-B hepatitis virus; NANBHV; HCV; antigen; detection;
KW diagnosis; antibodies.
XX
XX Hepatitis C virus.
XX
XX Key Location/Qualifiers
FH Misc-difference 1..122
FT /label= antigen
FT /note= "C22; R90936"
FT Misc-difference 199..328
FT /label= antigen
FT /note= "S2; R90935"
FT Misc-difference 1192..1457
FT /label= antigen
FT /note= "C33c; R90932"
FT Misc-difference 1569..1931
FT /label= antigen
FT /note= "C100; R90933"
FT Misc-difference 2054..2464
FT /label= antigen
FT /note= "NS5; R90934"
XX
XX EP693687-A1.
XX
XX 24-JAN-1996.
XX
XX 03-APR-1991; 91EP-0114016.
XX
XX 04-APR-1990; 90US-0504352.
XX (CHIR ) CHIRON CORP.
XX
XX Choo Q, Houghton M, Kuo G;
XX WPI; 1996-117956/13.
XX N-PSDB; T12710.
XX
XX Combinations of synthetic Hepatitis C Virus antigens - provide more
XX effective diagnosis of Non-A, Non-B Hepatitis
XX
XX Disclosure; Fig 1(A-Y); 53pp; English.
XX
XX The combination comprises an HCV antigen from the C domain (pref.
XX C22 - R90936) and at least one HCV antigen from the NS3 (pref. C33c
XX - R90932), NS4 (pref. C100 - R90933), S (pref. S2 - R90935) or NS5
XX (R90934) domain.
XX The antigens may in the form of a fusion protein, a simple physical
XX mixture, or the individual antigens commonly bound to a solid matrix.
XX They are pref. prepd. by recombinant DNA techniques (primers are
XX given in T12711-T12716), but can be synthesised or isolated from
XX HCV using affinity chromatography.
XX
XX Sequence 3011 AA;
XX
XX Query Match 100.0%; Score 1494; DB 17; Length 3011;
XX Best Local Similarity 100.0%; Pred. No. 8.3e-124;
XX Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 GAKQNVQLINTNGSHLNSALNCNDSLNTGWLGLFYHHKFNSSGCCPERLASCRPLTDF 60
XX 406 gakqnvqlintngshlntgwlglfyyhhkfnssgpcrpltdf 465
XX
XX 61 DQGWPIYANGSGDQRPYCHWHYPPKCGIYPAKSVCPYCFTPSPVYVCTTDRSGAP 120
XX 466 dqgwpisyangsgdqrpychwhypkpcgilypaksvcgpycftpspvvgttdrsgap 525
XX
XX 121 TYSWGENDDFVLNTRPPLGNWFGCTWMNSTGTFTKVCGAPPCVIGAGNNTLHCPTDC 180
XX 526 tyswgenddcvflntrpplgnwfgctwmnstgtftkvcgappcvigagntnlhcptdc 585
XX
XX 181 FRKHDPATYSRCGSGFWITPRCLVDYPRVRLWHYPCNTINTIPKIRMYVGGVEHRLAECN 240

```

```

Db 586 frkhpdatysrcgsgfwitprclvdypyrwhypctyntifkirmvvgvghrleaecn 645
QY 241 WTRGERCDLEDRDR 254
Db 646 wtrgercdledrdr 659

RESULT 9
W34480
ID W34480 standard; Protein; 3011 AA.
XX
XX W34480;
XX
XX 16-MAR-1998 (first entry)
XX
XX HCV polyprotein.
XX
XX PCR primer; amplify; HCV; hepatitis c virus; antigen combination; NS3;
XX C domain; S domain; NS5; HCV polyprotein; anti-HCV antibody; detection;
XX NS4.
XX
XX Hepatitis C virus.
XX
XX Key Location/Qualifiers
FH Misc-difference 366 /note= "can optionally be Arg"
FT
FT Misc-difference 372 /note= "can optionally be Thr"
FT
FT Misc-difference 867 /note= "can optionally be Thr"
FT
FT Misc-difference 1341 /note= "can optionally be Val"
FT
FT Misc-difference 2148 /note= "can optionally be Ile"
FT
FT Misc-difference 2883 /note= "can optionally be Asn"
FT
FT Misc-difference 3681 /note= "can optionally be Ser"
FT
FT Misc-difference 3690 /note= "can optionally be Thr"
FT
FT Misc-difference 4167 /note= "can optionally be Leu"
FT
FT Misc-difference 4323 /note= "can optionally be Val"
FT
FT Misc-difference 4701 /note= "can optionally be Tyr"
FT
FT Misc-difference 4752 /note= "can optionally be Ser"
FT
FT Misc-difference 5970 /note= "can optionally be Gly"
FT
FT Misc-difference 6183 /note= "can optionally be His"
FT
FT Misc-difference 6186 /note= "can optionally be Cys"
FT
FT Misc-difference 6402 /note= "can optionally be Val"
FT
FT Misc-difference 7386 /note= "can optionally be Ser"
FT
FT Misc-difference 7494 /note= "can optionally be Phe"
FT
FT Misc-difference 7497 /note= "can optionally be Ala"
FT
FT Misc-difference 7845 /note= "can optionally be Phe"
FT
FT Misc-difference 8409 /note= "can optionally be Gly"
FT
FT Misc-difference 9102 /note= "can optionally be Gly"
FT
FT Misc-difference 9327 /note= "can optionally be Pro"
XX
XX US5683864-A.

```

```

XX PD 04-NOV-1997.
XX PF 07-JUL-1992; 92US-0910760.
XX PR 07-JUL-1992; 92US-0910760.
PR 18-NOV-1987; 87US-0122714.
PR 30-DEC-1987; 87US-0139886.
PR 26-FEB-1988; 88US-0161072.
PR 06-MAY-1988; 88US-0191263.
PR 26-OCT-1988; 88US-0263584.
PR 14-NOV-1988; 88US-0271450.
PR 17-MAR-1989; 89US-0325338.
PR 20-APR-1989; 89US-0341334.
PR 21-APR-1989; 89US-0353896.
PR 04-APR-1990; 90US-0504332.
XX PA (CHIR ) CHIRON CORP.
XX PI Choo Q, Houghton M, Kuo G;
XX DR WPI: 1997-548976/50.
XX DR N-PSDB; T99981.
XX PT Combination of three hepatitis C virus antigens - used for detection
XX PS of specific antibodies to diagnose infection
XX PS Disclosure; Column 25-46; 57pp; English.
XX CC This sequence represents the Hepatitis C virus polyprotein. Fragments of
CC the DNA encoding this sequence can be amplified and used in the
CC combination of HCV antigens of the invention. The HCV antigen combination
CC comprises an antigen (Ag1) comprising the C domain (i.e. amino acids (aa)
CC 1-120 of the HCV polyprotein), or its immunologically reactive fragment
CC containing at least 8 aa. It also comprises two additional reactive fragments
CC two different polyprotein domains, including at least 8 aa from the NS3,
CC NS4, S or NS5 domains of the polyprotein, corresponding, respectively, to
CC aa 1050-1640; 1640-2000; 120-400 and 2000-3011 of the HCV polyprotein.
CC Alternatively, Ag1 contains at least 8 aa from the 1-122 or 9-177 aa
CC regions of the HCV polyprotein. These antigen combinations are used
CC diagnostically to detect anti-HCV antibodies, using any standard
CC immunoassay format. These antigen combinations have a broader range of
CC reactivity with antibodies than any antigen individually.
XX SQ Sequence 3011 AA;

Query Match 100.0%; Score 1494; DB 18; Length 3011;
Best Local Similarity 100.0%; Pred. NO. 8.3e-124;
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAKQNVQLINTGSHLNSTALNCNDLSLNTGLAGLFYHHKFNSSGCPERLASCRPLTDF 60
Db 406 gAkqnvqlIntgshlnstalnCNdSLntgWlaGlfYhHkfnssGcperlasCrlptdf 465
QY 61 DQGWGPISTVANGSGPDQRPYCHYHPKPCGIVPAKSVCGPVYCFPTSPVVGTTDRSGAP 120
Db 466 dGwgGpIstVangsgpDqrpYchYhPkpcgIvpaksvCgpVycfptspVvgTtdrsgap 525
QY 121 TYSWGENDDTVFLNTRPLGNWFECTWMNSTGFTKVCAGPCVJGGAGNNTLHCPTDC 180
Db 526 tYswgEnDdtvFlntrPlgnWfctwmnStgftkvcagpCvjggagNntlhcptdc 585
QY 181 FRKHDPATYSRCSGSPWITPRCLVDYFYRLWHYPCTINVTIFKIRMYGVGVEHRLAEACN 240
Db 586 frKhdpAtysrCsGspwItprClvDyfyRlwhYpctInvtIfkIrMygvGveHrleAacN 645
QY 241 WTRGERCDLEDRDR 254
Db 646 wtrGerCdleDrdr 659

RESULT 10

```

```

W40038
ID W40038 standard; Protein; 3011 AA.
XX AC
XX W40038;
XX DT 26-MAY-1998 (first entry)
XX DE HCV polyprotein.
XX KW Hepatitis C virus C domain; HCV; C antigen; immunological activity;
XX KW NS3 domain; NS4 domain; S domain; NS5 domain.
XX OS Hepatitis C virus.
XX FH Key
FH Domain 1..120
FT Location/Qualifiers
FT /label= C_domain
FT Modified-site 9
FT /note= "As given in the specification this amino
FT acid can also be Arg"
FT Modified-site 11
FT /note= "As given in the specification this amino
FT acid can also be Thr"
FT Domain 120..400
FT /label= S_domain
FT Modified-site 174
FT /note= "As given in the specification this amino
FT acid can also be Thr"
FT Modified-site 334
FT /note= "As given in the specification this amino
FT acid can also be Val"
FT Modified-site 603
FT /note= "As given in the specification this amino
FT acid can also be Ile"
FT Modified-site 847
FT /note= "As given in the specification this amino
FT acid can also be Asn"
FT Domain 1050..1640
FT /label= NS3_domain
FT Modified-site 1114
FT /note= "As given in the specification this amino
FT acid can also be Ser"
FT Modified-site 1217
FT /note= "As given in the specification this amino
FT acid can also be Thr"
FT Modified-site 1276
FT /note= "As given in the specification this amino
FT acid can also be Leu"
FT Modified-site 1328
FT /note= "As given in the specification this amino
FT acid can also be Val"
FT Modified-site 1452
FT /note= "As given in the specification this amino
FT acid can also be Tyr"
FT Modified-site 1472
FT /note= "As given in the specification this amino
FT acid can also be Ser"
FT Domain 1640..2000
FT Modified-site 1877
FT /label= NS4_domain
FT /note= "As given in the specification this amino
FT acid can also be Gly"
FT Modified-site 1948
FT /note= "As given in the specification this amino
FT acid can also be His"
FT Modified-site 1949
FT /note= "As given in the specification this amino
FT acid can also be Cys"
FT Domain 2000..3011
FT /label= NS5_domain
FT Modified-site 2021
FT /note= "As given in the specification this amino
FT acid can also be Val"

```

FT Modified-site 2348 /note= "As given in the specification this amino
FT acid can also be Ser"
FT Modified-site 2385 /note= "As given in the specification this amino
FT acid can also be Phe"
FT Modified-site 2386 /note= "As given in the specification this amino
FT acid can also be Ala"
FT Modified-site 2502 /note= "As given in the specification this amino
FT acid can also be Phe"
FT Modified-site 2690 /note= "As given in the specification this amino
FT acid can also be Gly"
FT Modified-site 2921 /note= "As given in the specification this amino
FT acid can also be Gly"
FT Modified-site 2996 /note= "As given in the specification this amino
FT acid can also be Pro"
FT FT XX
PN US5712087-A.
XX
PD 27-JAN-1998.
XX
XX 12-MAY-1995; 95US-0440519.
XX 07-JUL-1992; 92US-0910760.
PR 04-APR-1990; 90US-0504352.
PR 12-MAY-1995; 95US-0440519.
XX
XX (CHIR) CHIRON CORP.
PA
XX Choo Q, Houghton M, Kuo G;
PI
XX WPI: 1998-119973/11.
DR N-PSDB; V09989.
DR
XX Immunoassays for hepatitis C virus antibodies - using combinations
PT of antigenic fragments of HCV polyprotein
PT
XX Disclosure; Fig 1; 59pp; English.
XX
CC This sequence represents the hepatitis C virus (HCV) polyprotein which
CC is used in the construction of novel combinations of HCV antigens that
CC have a broader range of immunological activity than any single HCV
CC antigen. An example of such an antigen given in this specification
CC comprises a first antigen containing at least 8 amino acids of the
CC C domain of the HCV polyprotein and a second antigen comprising at least
CC 8 amino acids of the NS3 domain, the NS4 domain, the S domain or the NS5
CC domain of the HCV polyprotein in the form of a fusion protein, a physical
CC mixture or bound to a solid matrix.
CC Note: The features given in the specification as represented in the
CC feature table of W40038 differ from the positions indicated in Figure 1.
XX
XX Sequence 3011 AA:

Query Match 100.0%; Score 1494; DB 19; Length 3011;
Best Local Similarity 100.0%; Pred. NO. 8.3e-124; Indels 0; Gaps 0;
Matches 254; Conservative 0; Mismatches 0;
QY 1 GAKONVOLINTGSHLNSLTALNCNDLSLNTGLAGLFYHHKFNSSGCPERLASCRPLTDF 60
DB 406 gakqnvqlintgshlntalncndslntglaglfyhhkfnssgcperscrpltdf 465
QY 61 DQGMGPTSYANGSGDQRPQWHPKPCGIVPAKSVCGPVICFTSPVVGTTDRSGAP 120
DB 466 dqgwgplisyangsgdpqrpqwyhpkpcgivpaksvcgpcvftspvvgtttdrsgap 525
QY 121 TYSXGENDTDVFLNTRPPLGNWFGCTWMNSTGFTKVCAPPCTVGGAGNLTLCPTQC 180
TYSXGENDTDVFLNTRPPLGNWFGCTWMNSTGFTKVCAPPCTVGGAGNLTLCPTQC 180

DB 526 tyswgendtdvflntrpplgnwfgctwmnstgftkvcappcvvggagntllhcptdc 585
QY 181 FRKHDPATYSRCGSGPWITPRCLVDYPYRLWHYPCPTINTIFKIRMYVGVGVEHRLAAGN 240
DB 586 frkhdpatsrcsgsgpwitprclvdypyrwlwhypctintyifkirmvvgvvehrleaacn 645
QY 241 WTRGERCDLEDRDR 254
DB 646 wtrgercdledrdr 659
RESULT 11
R31621
ID R31621 standard; protein; 3011 AA.
XX
AC R31621;
XX
DT 24-MAY-1993 (first entry)
XX
DE Hepatitis C virus (HCV) polyprotein.
XX
DE Hepatitis; liver disease; HCV1; monoclonal antibody; epitope;
KW immobilised reagent; immunoassay; diagnosis; detection; treatment;
KW infection.
KW
XX
OS Hepatitis C virus type 1.
XX
FH Location/Qualifiers
FT 1..191
FT Domain /label= C domain
FT /note= "nucleocapsid protein"
FT Domain 192..383
FT /label= E1
FT /note= "virion envelope protein"
FT Domain 384..800
FT /label= E2/NS1
FT /note= "possible envelope"
FT Domain 800..1050
FT /label= NS2
FT /note= "unknown function"
FT Domain 1050..1650
FT /label= NS3
FT /note= "putative protease domain"
FT Domain 1651..2100
FT /label= NS4
FT /note= "unknown function"
FT Domain 2100..3011
FT /label= NS5
FT /note= "polymerase"
PN WO9300365-A.
XX
XX 07-JAN-1993.
PD
XX 24-JUN-1992; 92WO-US05388.
XX
XX 24-JUN-1991; 91US-0722489.
XX
XX (CHIR) CHIRON CORP.
XX
XX Chien DY, Rutter W;
XX
XX WPI: 1993-036334/04.
XX
XX Polypeptide(s) comprising truncated hepatitis C virus sequences -
PT for detection, prevention and treatment of hepatitis C infection
XX
XX Claim 1; Fig 1; 80pp; English.
PS
XX This sequence represents the polyprotein of the HCV prototype
CC isolate HCV1. When compared with all known viral sequences, small
CC but significant co-linear homologies are observed with the non-
CC structural proteins of the flavivirus family, and with the

CC pestiviruses. The domains shown in the features table are however,
CC tentatively assigned. The polyprotein, and epitopes of it are
CC useful for inducing immunological response in a subject against
CC HCV. The presence of Abs against HCV can be detected using an
CC immunoassay.
XX
SQ Sequence 3011 AA;

Query Match 99.9%; Score 1492; DB 14; Length 3011;
Best Local Similarity 99.6%; Pred. No. 1.3e-123;
Matches 253; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAKQNVQLINTNGSWHLNSTALNCNDSLNTGWLGLAGFYHHKFNSSGCPERLASCRPLTDF 60
DB 406 gakqnvqlintngswhlNSTALNCNDSLNTGWLGLAGFYHHKFNSSGCPERLASCRPLTDF 465
QY 61 DQGWGPISYANGSGDPQRPYCHWHYPPKPCGIVPAKSVCGPYVCFPTSPVVGTTDRSGAP 120
DB 466 dqgwgplsyangsgdpqrpYCHWHYPPKPCGIVPAKSVCGPYVCFPTSPVVGTTDRSGAP 525
QY 121 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCAPPVCVIGGAGNNTLHCPTDC 180
DB 526 tyswgentdtdfVLNTRPPLGNWFGCTWMNSTGFTKVCAPPVCVIGGAGNNTLHCPTDC 585
QY 181 FRKHDPATYSRCGSGPWINTPRCLVDYPYRLWHYPCPTINTYFIKIRMYVGGVHRLEAACN 240
DB 586 frkhdpatysrcsgsgpwintprclvdypYRLWHYPCPTINTYFIKIRMYVGGVHRLEAACN 645
QY 241 WTRGERCDLEDRDR 254
DB 646 wtrgercdleddrdr 659

RESULT 12
R08124
ID R08124 standard; protein; 2955 AA.
AC R08124;
XX
DT 23-JAN-1991 (first entry)
XX
DE Hepatitis C virus putative polyprotein.
XX
KW Hepatitis C virus (HCV); antiviral agent.
XX
OS Hepatitis C virus.
XX
FH Key Location/Qualifiers
FT Misc-difference 9..9 /label=K or R
FT Misc-difference 11..11 /label=N or T
FT Misc-difference 176..176 /label=I or T
FT Misc-difference 334..334 /label=M or V
FT Misc-difference 603..603 /label=I or L
FT Misc-difference 848..848 /label=Y or N
FT Misc-difference 1114..1114 /label=P or S
FT Misc-difference 1117..1117 /label=S or T
FT Misc-difference 1276..1276 /label=P or L
FT Misc-difference 1454..1454 /label=C or Y
FT Misc-difference 1471..1471 /label=T or S
FT Misc-difference 1877..1877 /label=E or G

FT Misc-difference 1948..1948 /label=L or H
FT Misc-difference 1949..1949 /label=S or C
FT Misc-difference 2021..2021 /label=V or G
FT Misc-difference 2349..2349 /label=T or S
FT Misc-difference 2385..2385 /label=Y or F
FT Misc-difference 2386..2386 /label=S or A
FT Misc-difference 2502..2502 /label=L or F
FT Misc-difference 2690..2690 /label=R or G
FT Misc-difference 2921..2921 /label=R or G
XX
PN EP388232-A.
XX
PD 19-SEP-1990.
XX
XX
PF 16-MAR-1990; 90EP-0302866.
XX
PR 18-MAY-1989; 89US-0355002.
PR 17-MAR-1989; 89US-0325338.
PR 20-APR-1989; 89US-0341334.
XX
PA (CHIR-) CHIRON CORP.
XX
XX
PI Houghton M, Choo QL, Kuo G;
DR WPI; 1990-284418/38.
DR N-PSDB; Q05956.
XX
PT Hepatitis C virus DNA - used for producing probes,
PT polypeptide(s), antibodies and anti-sense polynucleotide(s) for
PT diagnosis and therapy.
XX
PS Disclosure: Fig 17; 83pp; English.
XX
CC HCV cDNA libraries were constructed using pooled serum from a
CC chimpanzee with chronic HCV infection. A lambda gt10 library was
CC screened with probes derived from previously isolated clones. The
CC ORF is derived from the overlapping clones b114a, ag30a, CA205a,
CC CA290a, CA216a, p14a, CA167b, CA156e, CA84a, CA59a, K9-1, 261, 131,
CC 12f, 141, 11b, 7f, 8h, 33c, 40b, 37b, 35, 36, 81, 32, 33b, 25c,
CC 14c, 8f, 33f, 33g, 39c, 35f, 19g, 26g, 15e, b5a and 16jh.
CC Polypeptide encoded by this sequence can be used to design probes
CC for the detection of HCV nucleic acids, in screening programmes
CC for antiviral agents and in preparing blood free of HCV. The
CC sequence contains 188 (overlapping) peptides which are claimed as
CC HCV epitopes.
CC See also Q05955.
XX
SQ Sequence 2955 AA;

Query Match 99.7%; Score 1489; DB 11; Length 2955;
Best Local Similarity 99.6%; Pred. No. 2.3e-123;
Matches 253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAKQNVQLINTNGSWHLNSTALNCNDSLNTGWLGLAGFYHHKFNSSGCPERLASCRPLTDF 60
DB 406 gakqnvqlintngswhlNSTALNCNDSLNTGWLGLAGFYHHKFNSSGCPERLASCRPLTDF 465
QY 61 DQGWGPISYANGSGDPQRPYCHWHYPPKPCGIVPAKSVCGPYVCFPTSPVVGTTDRSGAP 120
DB 466 dqgwgplsyangsgdpqrpYCHWHYPPKPCGIVPAKSVCGPYVCFPTSPVVGTTDRSGAP 525
QY 121 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCAPPVCVIGGAGNNTLHCPTDC 180

Db 526 tyswgendtdvfvlntrppignwfgctwmnstgftkvcgappcvigagnttlhcptdc 585

Qy 181 FRKHDPATYSRCGSGPWITPRCLVDYPRYLWHYPCTINYTIKIRMYVGGVEHRELAACN 240
 |||||
 Db 586 frkhdpatysrcsgsgpwitprclvdypylwhypctinytifclrmvvgvghrleaach 645
 |||||

Qy 241 WTRGERCDLEDRDR 254
 |||||
 Db 646 wtrgercdleldrdr 659

RESULT 13

R34009
 ID R34009 standard; Protein; 2816 AA.
 XX
 AC R34009;
 XX
 DT 26-JUL-1993 (first entry)
 XX
 DE HCV-1 polyprotein.
 DE
 XX
 KW Polymerase chain reaction; PCR; amplify; primer; hepatitis C virus;
 KW HCV; asymptomatic; chronically infected; epitope; viral isolate;
 KW domain; immunological; cross-reactive; envelope protein; vaccine;
 KW gp53(BVDV)/gp55; hog cholera virus; pestivirus; NS1; flavivirus.
 XX
 OS Hepatitis C virus.
 XX
 FN WO9306126-A.
 XX
 PD 01-APR-1993.
 XX
 PF 11-SEP-1992; 92WO-US07683.
 XX
 PR 13-SEP-1991; 91US-0759575.
 XX
 PA (CHIR) CHIRON CORP.
 XX
 PI Houghton M, Weiner AJ;
 XX
 DR WPI; 1993-117468/14.
 XX
 PT Immuno-reactive hepatitis C virus polypeptide compsns. - contg.
 PT at least 2 sequences from the first variable domain of distinct
 PT HCV isolates
 XX
 PS Disclosure; Fig 9; 106pp; English.
 XX
 CC This sequence represents the entire hepatitis C virus polyprotein.
 CC HCV is a member of the flavivirus family and appears to encode a basic
 CC polypeptide domain ("C") at the N-terminal of the viral polyprotein,
 CC followed by two glycoprotein domains ("E1", "E2/NS1"), upstream of the
 CC nonstructural genes NS2 through NS5. See also Q39134-48, R33982-
 CC 4008 and R38088-89.
 CC
 XX
 SQ Sequence 2816 AA;

Query Match 99.5%; Score 1486; DB 14; Length 2816;
 Best Local Similarity 99.6%; Pred. No. 3.9e-123;
 Matches 253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAKQNVQLINTNGSWHLNSTALNCNDSINTGLAGLFYHHKFNSSGCGPERLASCRPLTDF 60
 |||||
 Db 406 gakqnvqlintngswhlNSTALNCNDSINTGLAGLFYHHKFNSSGCGPERLASCRPLTDF 465

Qy 61 DQGWGPISYANGSGDPDRPYCWHYPPKPCGIVPAKSVCGPVYCFPTSPVVGTTDRSGAP 120
 |||||
 Db 466 dggwgpisyangsgdpdrpycwhypkpcgivpaksvcgpyvcftspvvgtttdrsagp 525

Qy 121 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCGAPPCVIGAGNNTLHCPTDC 180
 |||||
 Db 526 tyswgendtdvfvlntrppignwfgctwmnstgftkvcgappcvigagnttlhcptdc 585

Qy 181 FRKHDPATYSRCGSGPWITPRCLVDYPRYLWHYPCTINYTIKIRMYVGGVEHRELAACN 240
 |||||
 Db 586 frkhdpatysrcsgsgpwitprclvdypylwhypctinytifclrmvvgvghrleaach 645
 |||||

Qy 241 WTRGERCDLEDRDR 254
 |||||
 Db 646 wtrgercdleldrdr 659

RESULT 14

R33185
 ID R33185 standard; peptide; 621 AA.
 XX
 AC R33185;
 XX
 DT 03-JUL-1993 (first entry)
 XX
 DE Sequence of subfragment pHCV80 (AAs 365-731) of the hepatitis
 DE C virus (HCV) genome.
 DE
 XX
 KW Immunogenic peptide; hepatitis C virus; immunogenic domain;
 KW monoclonal antibody; diagnosis; detection; therapy.
 XX
 OS Synthetic.
 XX
 FN WO9304205-A.
 XX
 PD 04-MAR-1993.
 XX
 PF 21-AUG-1992; 92WO-US07189.
 XX
 PR 21-AUG-1991; 91US-0748292.
 XX
 PA (ABBO) ABBOTT LAB.
 XX
 PI Dailey SH, Desai SM, Devare SG, Johnson JE, Mehta SU;
 XX
 DR WPI; 1993-094030/11.
 XX
 PT Monoclonal antibodies specific for hepatitis C virus E2-NS1
 PT antigen - useful for diagnosis and evaluation of HCV infections
 PT and in differentiation studies
 XX
 PS Example; Pages 31-34; 48pp; English.
 XX
 CC Immunogenic domains of E2/NS1 region of HCV genome encompassing AAs
 CC 600-720 were mapped with PEPSCAN analysis. Based on the EIA reactivity
 CC of a panel of HCV positive sera, peptide R33184 was chosen as the
 CC immunogen for the generation of monoclonal antibodies to HCV NS1.
 CC Several individual oligos representing AAs 365-731 of HCV genome
 CC were ligated and cloned as three separate EcoRI/BamHI subfragments
 CC into the CKS fusion vector pJ0200. The sequences of these three
 CC subfragments - pHCV80, pHCV77 and pHCV65 are in R33185-7.
 CC Analysis showed that MAbS H13C113 and p23C163 showed reactivity
 CC pHCV80 and pHCV65.
 XX
 SQ Sequence 621 AA;

Query Match 98.6%; Score 1473; DB 14; Length 621;
 Best Local Similarity 97.3%; Pred. No. 9.5e-123;
 Matches 253; Conservative 0; Mismatches 1; Indels 6; Gaps 1;

Qy 1 GAKQNVQLINTNGSWHLNSTALNCNDSINTGLAGLFYHHKFNSSGCGPERLASCRPLTDF 60
 |||||
 Db 290 gakqnvqlintngswhlNSTALNCNDSINTGLAGLFYHHKFNSSGCGPERLASCRPLTDF 349

Qy 61 DQGWGPISYANGSGDPDRPYCWHYPPKPCGIVPAKSVCGPVYCFPTSPVVGTTDRSGAP 120
 |||||
 Db 350 dggwgpisyangsgdpdrpycwhypkpcgivpaksvcgpyvcftspvvgtttdrsagp 409

Qy 121 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCGA-----PPCVIGAGNNTL 174

Db 410 tyswgentdvfvlnttrpplgnwfgctwmnstgftkvcgappcvlgppcviggagntll 469
Qy 175 HCPDCEFRKHPDATYSCGSGPWITPRCLVDYPRRLWHYPCTINITYTIFKIRMYVGGVEHR 234
Db 470 hcptdcfrkhpdatyscgsgpwitprclvdpyrlwhypctinytlfkirmvvgvgehr 529
Qy 235 LEACNWTGRGCDLEDRDR 254
Db 530 leaecnwtgrgcdledrdr 549

RESULT 15

ID R33591 standard; protein; 622 AA.
AC R33591;
DT 05-JUL-1993 (first entry)
DE HCV CKS-NS1S1-NS1S2 fusion antigen.
DE Hepatitis C Virus; non-A, non-B hepatitis virus; NANBH;
KW non-structural protein; CMP-KDO synthetase; CKS fusion protein;
KW CTP:OMP-3-deoxy-manno-octulosonate cytidyl transferase;
KW immunoassay; pHCV-80.
OS Hepatitis C Virus.

Key Location/Qualifiers
FT Misc-difference 622
FT /note= "not defined"

PN W09304088-A.
PD 04-MAR-1993.
XX 21-AUG-1992; 92WO-US07188.
XX 21-AUG-1991; 91US-0748561.
PA (ABBO) ABBOTT LAB.
PI Dailey SH, Desai SM, Devare SG;
DR WPI; 1993-093941/11.
XX Hepatitis C assay using recombinant NS1 region antigens - for
PT detecting-antibodies and antigen in body fluids from individuals
PT exposed to hepatitis C virus

PS Claim 4; Page 44-46; 175pp; English.
XX The construction of pHCV-80 (NS1S1-NS1S2) involved using the
CC SacI/BamHI insert from pHCV-65 (see R33589) and ligating that into
CC the SacI/BamHI vector backbone of pHCV-77 (see R33588). The resultant
CC HCV gene represents amino acids 365-731 of the HCV genome. This
CC resulted in a 1101bp EcoRI/BamHI fragment of HCV cloned into the
CC CKS fusion vector pJ0200. The complete amino acid sequence of the
CC antigen is designated pHCV-80 (i.e. R33591). The resultant fusion
CC protein HCV CKS NS1S1-NS1S2 consists of 239 amino acids of CKS,
CC seven amino acids contributed by linker DNA sequences and 367 amino
CC acids from the NS1 region of the HCV genome. The fusion protein is
CC used to detect antibodies and antigen in body fluids from
CC individuals exposed to HCV.

XX Sequence 622 AA;

Query Match 98.6%; Score 1473; DB 14; Length 622;
Best Local Similarity 97.3%; Pred. No. 9.5e-123;
Matches 253; Conservative 0; Mismatches 1; Indels 6; Gaps 1;

Qy 1 GAKQNVQLINTNGSWHLNSTALNCNDSLNTGWLAGLFYHHKFNSSGCPERLASCRPI.TDF 60
Db 290 gakqnvqlintngswhlntalnncndsintgwlaglffyhhkfnssgcpelascrpltdf 349
Qy 61 DOGWGPISYANGSGPDORPYCWHYPPKPCGIVPAKSVCGPVYCFTPSPVVGTTDRSGAP 120
Db 350 dqgwqqlsyangsgpdqrpqcywhypkpcgilypaksvcgvpycftpspvvvgttcdrsgap 409
Qy 121 TYSWGENDDTVFVLNNTTRPPLGNMFGCTWMNSTGFTKVCGA-----PPCVIGGAGNNTL 174
Db 410 tyswgentdvfvlnttrpplgnwfgctwmnstgftkvcgappcvlgppcviggagntll 469
Qy 175 HCPDCEFRKHPDATYSCGSGPWITPRCLVDYPRRLWHYPCTINITYTIFKIRMYVGGVEHR 234
Db 470 hcptdcfrkhpdatyscgsgpwitprclvdpyrlwhypctinytlfkirmvvgvgehr 529
Qy 235 LEACNWTGRGCDLEDRDR 254
Db 530 leaecnwtgrgcdledrdr 549

Search completed: March 6, 2001, 11:51:51
Job time: 74 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run On: March 6, 2001, 11:53:15 ; Search time 45.68 Seconds
(without alignments)
377.556 Million cell updates/sec

Title: US-09-407-430-3

Perfect score: 1494

Sequence: 1 GAKQNVQLINTNGSHLNTST.....LEAACNWTGRGCDLEDRDR 254

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_66: *
1: pirl: *
2: pirl: *
3: pirl: *
4: pirl: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1494	100.0	3011	1 GNVVC3	genome polyprotein
2	1438	96.3	3011	1 S40770	genome polyprotein
3	1398	93.6	3011	1 GNVVCH	genome polyprotein
4	1367	91.5	716	2 JQ1366	polyprotein - hepa
5	1273	85.2	3010	1 GNVVTV	genome polyprotein
6	1272	85.1	640	2 JQ1584	genome polyprotein
7	1254	83.9	3010	1 GNVVTC	genome polyprotein
8	1253	83.9	3010	1 GNVVCT	genome polyprotein
9	1244	83.3	782	2 S19875	genome polyprotein
10	1242.5	83.2	3014	1 JC5620	genome polyprotein
11	1236	82.7	782	2 S19876	genome polyprotein
12	1234	82.6	3010	1 A45573	genome polyprotein
13	1234	82.6	3010	1 S18030	genome polyprotein
14	1228	82.2	782	2 S18032	genome polyprotein
15	1199	80.3	782	2 S18031	genome polyprotein
16	1171	78.4	787	2 PNU677	genome polyprotein
17	1170	78.3	3033	1 GNVVJB	genome polyprotein
18	1156	77.4	3033	1 JQ1303	genome polyprotein
19	1122	75.1	350	2 S35631	genome polyprotein
20	942	63.1	234	2 S32742	genome polyprotein
21	936	62.7	235	2 S32747	genome polyprotein
22	905	60.6	237	2 S32744	genome polyprotein
23	756	50.6	415	2 PC3407	genome polyprotein
24	649.5	43.5	876	2 PC3219	polyprotein - hepa
25	597	40.0	513	2 A44150	structural protein
26	550	36.8	138	2 S24080	envelope protein -
27	539	36.1	138	2 S24081	envelope protein -
28	522	34.9	523	2 JQ1926	polyprotein - hepa
29	521	34.9	138	2 S24074	envelope protein -

ALIGNMENTS

RESULT 1

GNVVC3

genome polyprotein - hepatitis C virus (strain HCV-1)

N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstru
-protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 26-May-2000

C:Accession: A39166; PQ0403; PQ0404

R:Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.;

Proc. Natl. Acad. Sci. U.S.A. 88, 1131-1141, 1991

A:Title: Genetic organization and diversity of the hepatitis C virus.

A:Reference number: A39166; MUID:91172826

A:Accession: A39166

A:Molecule type: mRNA

A:Residues: 1-3011 <CHO>

A:Cross-references: GB:M62321; NID:9329873; PIDN:AAA45676.1; PID:9329874

R:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Pollett, E.; Yap,

J. Gen. Virol. 73, 1131-1141, 1992

A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship t

A:Reference number: PQ0393; MUID:92268871

A:Accession: PQ0403

A:Molecule type: genomic RNA

A:Residues: 1577-1633 <CHA>

A:Cross-references: DBJ:D10128

A:Experimental source: isolates E-b16

A:Accession: PQ0404

A>Status: preliminary

A:Molecule type: genomic RNA

A:Residues: 1577-1633 <CH2>

A:Experimental source: isolates E-b17

C:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology

C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruct

F:1-115/Product: capsid protein C #status predicted <CPC>

F:192-389/Product: major envelope protein M #status predicted <EPM>

F:390-729/Product: nonstructural protein NS1 #status predicted <MEP>

F:730-1006/Product: nonstructural protein NS2 #status predicted <NS1>

F:1007-1615/Product: hepacivirin #status predicted <NS3>

F:1230-1493/Domain: DEAD/H box helicase homology <DEAD>

F:1230-1237/Region: nucleotide-binding motif A (P-loop)

F:1312-1317/Region: nucleotide-binding motif B

F:1316-1319/Region: DEXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>

F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>

F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077

Query Match

Best Local Similarity 100.0%; Score 1494; DB 1; Length 3011;

Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 586 FRKYPEATYSRGSGPRITPRCMVDYPIRLWHYPCTINTIFKVRMYVGGVEHRLAECN 645
 QY 241 WTRGERCDLEDRDR 254
 Db 646 WTRGERCDLEDRDR 659

RESULT 4
 JQ1366
 polyprotein - hepatitis C virus (French isolate) (fragments)
 C:Species: hepatitis C virus
 C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 29-May-1998
 C:Accession: JQ1366
 J:Krensdorf, D.; Porchon, C.; Kim, J.P.; Reyes, G.R.; Brechot, C.
 J:Gen. Virol. 72, 2557-2561, 1991
 A:Title: Partial nucleotide sequence analysis of a French hepatitis C virus: Implication
 A:Reference number: JQ1366; MUID:92013977
 A:Accession: JQ1366
 A:Molecule type: genomic RNA
 A:Residues: 1-716 <KRE>
 C:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology
 C:Keywords: glycoprotein; polyprotein
 F:84,90,97,115,143,199,223,243,290,312/Binding site: carbohydrate (Asn) (covalent) #stat

Query Match 91.5%; Score 1367; DB 2; Length 716;
 Best Local Similarity 87.4%; Pred. No. 7.6e-104;
 Matches 222; Conservative 25; Mismatches 74; Indels 0; Gaps 0;
 QY 1 GAKQNVQLINTNGSHLNTALNCNDLSLNTGWLGLFYHHKFNSSGCPERLASCRPLTDF 60
 Db 73 GAKQDIQLINTNGSHLNTALNCNESLDTGWAGLFYHHKFNSSGCPERMASCRPLADF 132
 QY 61 DQGWGPISYANGSGPDQRPYCHYPPKPGCIIVPAKSVCGPVYCFPTSPVVGTTDRSGAP 120
 Db 133 DQGWGPISYANGSGPEHRYCHYPPKPGCIIVPAQTVCGPVYCFPTSPVVGTTNKLGP 192
 QY 121 TYSWGENDTDFVLNTRPLGNWFCGTWNNSTGFTKVCAPPVCVIGGAGNNTLHCPTDC 180
 Db 193 TTNMGENDTDFVLNTRPLGNWFCGTWNNSTGFTKVCAPPVCVIGGAGNNTLYCPTDC 252
 QY 181 FRKHDPATYSRSGSGPWITPRCLVDYPIRLWHYPCTINTIFKIRMYVGGVEHRLAECN 240
 Db 253 FRKHPEATYSRSGSGPWITPRCLVGYPIRLWHYPCTVNTLTKVRMYVGGVEHRLQVACN 312
 QY 241 WTRGERCDLEDRDR 254
 Db 313 WTRGERCNLEDRDR 326

RESULT 5
 GNVYTW
 genome polyprotein - hepatitis C virus (strain Taiwan)
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 A:Note: host Homo sapiens (man)
 C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 26-May-2000
 C:Accession: A40244
 R:Chen, P.J.; Lin, M.H.; Tai, K.F.; Lin, P.C.; Lin, C.J.; Chen, D.S.
 Virology 188, 102-113, 1992
 A:Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the
 A:Reference number: A40244; MUID:92230206
 A:Accession: A40244
 A:Molecule type: genomic RNA
 A:Residues: 1-3010 <CHE>
 C:Cross-references: GB:M84754
 C:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural
 F:1-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <MEE>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>

F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: hepatitis C virus genome polyprotein; DEAD/H box helicase homology <NS3>
 F:1230-1493/Product: DEAD/H box helicase homology <DEAD>
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)
 F:1312-1317/Region: nucleotide-binding motif B
 F:1316-1319/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,233,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,204,

Query Match 85.2%; Score 1273; DB 1; Length 3010;
 Best Local Similarity 81.9%; Pred. No. 1.5e-95;
 Matches 208; Conservative 19; Mismatches 27; Indels 0; Gaps 0;
 QY 1 GAKQNVQLINTNGSHLNTALNCNDLSLNTGWLGLFYHHKFNSSGCPERLASCRPLTDF 60
 Db 406 GASQKIQLINTNGSHLNTALNCNDLSLNTGWLGLFYHHKFNSSGCPERMASCRSIDKF 465
 QY 61 DQGWGPISYANGSGPDQRPYCHYPPKPGCIIVPAKSVCGPVYCFPTSPVVGTTDRSGAP 120
 Db 466 DQGWGPITVTEADIQDORPYCHYPPKPGCIIVPAQTVCGPVYCFPTSPVVGTTDRFGAP 525
 QY 121 TYSWGENDTDFVLNTRPLGNWFCGTWNNSTGFTKVCAPPVCVIGGAGNNTLHCPTDC 180
 Db 526 TYSWGENETDVLNTRPLGNWFCGTWNNSTGFTKTCGGPPCNTGGGNNTLVCPDTC 585
 QY 181 FRKHDPATYSRSGSGPWITPRCLVDYPIRLWHYPCTINTIFKIRMYVGGVEHRLAECN 240
 Db 586 FRKHPEATYTKCGSPWLTPRCMVDYPIRLWHYPCTVNTLTKVRMYVGGVEHRLAECN 645
 QY 241 WTRGERCDLEDRDR 254
 Db 646 WTRGERCDLEDRDR 659

RESULT 6
 JQ1584
 genome polyprotein - hepatitis C virus (strain U.K.) (fragment)
 N:Contains: core protein C; envelope protein E1; envelope protein E2; nonstructural
 C:Species: hepatitis C virus
 C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 28-May-1999
 C:Accession: JQ1584
 R:Kumar, U.; Cheng, D.; Thomas, H.; Monjardino, J.
 J:Gen. Virol. 73, 1521-1525, 1992
 A:Title: Cloning and sequencing of the structural region and expression of putative c
 A:Reference number: JQ1584; MUID:92300349
 A:Accession: JQ1584
 A:Molecule type: genomic RNA
 A:Residues: 1-640 <KUM>
 A:Cross-references: GB:X84079; NID:G643119; PIDN:CAA58888.1; PID:G643120
 C:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology
 C:Keywords: core protein; envelope protein; glycoprotein; nonstructural protein; poly
 F:1-191/Product: core protein C #status predicted <CPC>
 F:192-389/Product: envelope protein E1 #status predicted <EEL>
 F:390-640/Product: envelope protein E2 and nonstructural protein NS1 #status predite
 F:196,209,234,305,417,430,448,476,540,556,576,623/Binding site: carbohydrate (Asn) (c

Query Match 85.1%; Score 1272; DB 2; Length 640;
 Best Local Similarity 88.5%; Pred. No. 3.7e-96;
 Matches 208; Conservative 17; Mismatches 10; Indels 0; Gaps 0;
 QY 1 GAKQNVQLINTNGSHLNTALNCNDLSLNTGWLGLFYHHKFNSSGCPERLASCRPLTDF 60
 Db 406 GSRQNTQLINTNGSHLNTALNCNLSLETGWLGLFYHHKFNSSGCTERMASCRPLADF 465
 QY 61 DQGWGPISYANGSGPDQRPYCHYPPKPGCIIVPAKSVCGPVYCFPTSPVVGTTDRSGAP 120
 Db 466 DQGWGPITVANGSGPEHRYCHYPPKPGCIIVPAQNVCGPVYCFPTSPVVGTTDRSGAP 525
 QY 121 TYSWGENDTDFVLNTRPLGNWFCGTWNNSTGFTKVCAPPVCVIGGAGNNTLHCPTDC 180

Db 526 TYNWGSNDTDYFVNLNTRPPLGNFGCTWNNSSGFTKVCAGPPCNGVGNNTLHCPTDC 585
QY 181 FRKHDPATYSCGSGPWTTPCLVDYPRYLWHYPCTTNTYTFKIRMYVGGVEHRL 235
Db 586 FRKHPEATYSCGSGPWTTPCLVDYPRYLWHYPCTTNTYTFKIRMYVGGVEHRL 640

RESULT 7
GNWVTC
genome polyprotein - hepatitis C virus
N;Contains: capsid protein C; envelope protein M; hepatitis C virus (strain J)
C;Species: hepatitis C virus
C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 26-May-2000
C;Accession: A38465
R;Takamizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E.;
J. Virol. 65, 1105-1113, 1991
A;Title: Structure and organization of the hepatitis C virus genome isolated from human
A;Reference number: A38465; MUID: 91140698
A;Accession: A38465
A;Molecule type: genomic RNA
A;Residues: 1-3010 <KAT>
A;Cross-references: EMBL: M58335; NID: g329770; PIDN: AAA72945.1; PID: g329771
C;Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS2 #status predicted <NS2>
F;730-1006/Product: nonstructural protein NS4a #status predicted <NS4a>
F;1007-1615/Product: hepatitis C virus genome polyprotein; DEAD/H box helicase homology
F;1230-1493/Domain: hepatitis C virus genome polyprotein; DEAD/H box helicase homology
F;1312-1317/Region: nucleotide-binding motif A (P-loop)
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F;196, 209, 234, 250, 305, 325, 417, 423, 430, 448, 532, 540, 556, 576, 623, 645, 1213, 1255, 2041, 2077, 22

Query Match 83.9%; Score 1254; DB 1; Length 3010;
Best Local Similarity 81.1%; Pred. No. 5.3e-94;
Matches 206; Conservative 16; Mismatches 32; Indels 0; Gaps 0;

QY 1 GAKNVQLINTNGSWHLNSTALNCNDSLNTGWLGLFHHKFNSSGCGPERLASCRPLTDF 60
Db 406 GPSOKIQLVNTNGSWHINRTALNCNDSLNTGFLAALFTHSFNSSGCGPERMAQCRTIDKF 465
QY 61 DOGWGPTISYANGSGDQRPYCHWYPPKPGIVPAKSGVCGPVYCTPSPVVGTTDRSGAP 120
Db 466 DOGWGPTIYAESSRSDQRPYCHWYPPQCTIVPAKSGVCGPVYCTPSPVVGTTDRSGAP 525
QY 121 TYSNGENDTDYFVNLNTRPPLGNFGCTWNNSSGFTKVCAGPPCNGVGNNTLHCPTDC 180
Db 526 TYRNGENETDVLNTRPPLGNFGCTWNNSSGFTKVCAGPPCNGVGNNTLHCPTDC 585
QY 181 FRKHDPATYSCGSGPWTTPCLVDYPRYLWHYPCTTNTYTFKIRMYVGGVEHRLAECN 240
Db 586 FRKHPEATYSCGSGPWTTPCLVDYPRYLWHYPCTTNTYTFKIRMYVGGVEHRLAECN 645
QY 241 WTRGERCDLEDRDR 254
Db 646 WTRGERCDLEDRDR 659

RESULT 8
GNWVTC
genome polyprotein - hepatitis C virus (strain J)
N;Contains: capsid protein C; envelope protein M; major envelope protein E; nonstructural
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus

C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 16-Jun-2000
C;Accession: A39253; PS0086
R;Kato, N.; Hijikata, M.; Ootsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Sugimura, T.; Shi
Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990
A;Title: Molecular cloning of the human hepatitis C virus genome from Japanese patient
A;Reference number: A39253; MUID: 91088550
A;Accession: A39253
A;Molecule type: genomic RNA
A;Residues: 1-3010 <KAT>
A;Cross-references: GB: D90208; NID: g221610; PIDN: BAA14233.1; PID: g221611
R;Kato, N.; Ohkoshi, S.; Shimotohno, K.
Proc. Jpn. Acad. 65B, 219-223, 1989
A;Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence v
A;Reference number: PS0085
A;Accession: PS0085
A;Molecule type: genomic RNA
A;Residues: 2650-2707 <KA2>
A;Experimental source: Japanese isolate
C;Comment: The cleavage sites of this polyprotein have not been determined.
C;Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology
C;Keywords: ATP; glycoprotein; hydrolase; P-loop; polyprotein; serine proteinase; tra
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepatitis C virus genome polyprotein; DEAD/H box helicase homology
F;1230-1493/Domain: hepatitis C virus genome polyprotein; DEAD/H box helicase homology
F;1312-1317/Region: nucleotide-binding motif A (P-loop)
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F;196, 209, 234, 250, 305, 325, 417, 423, 430, 448, 532, 556, 576, 623, 645, 1213, 1255, 2041, 2077, 224

Query Match 83.9%; Score 1253; DB 1; Length 3010;
Best Local Similarity 79.5%; Pred. No. 6.4e-94;
Matches 202; Conservative 25; Mismatches 27; Indels 0; Gaps 0;

QY 1 GAKNVQLINTNGSWHLNSTALNCNDSLNTGWLGLFHHKFNSSGCGPERLASCRPLTDF 60
Db 406 GPSOKIQLVNTNGSWHINRTALNCNDSLNTGFLAALFTHSFNSSGCGPERMAQCRTIDKF 465
QY 61 DOGWGPTISYANGSGDQRPYCHWYPPKPGIVPAKSGVCGPVYCTPSPVVGTTDRSGAP 120
Db 466 DOGWGPTIHDMPESDQRPYCHWYPPKPGIVPAKSGVCGPVYCTPSPVVGTTDRSGAP 525
QY 121 TYSNGENDTDYFVNLNTRPPLGNFGCTWNNSSGFTKVCAGPPCNGVGNNTLHCPTDC 180
Db 526 TYSNGENETDVLNTRPPLGNFGCTWNNSSGFTKVCAGPPCNGVGNNTLHCPTDC 585
QY 181 FRKHDPATYSCGSGPWTTPCLVDYPRYLWHYPCTTNTYTFKIRMYVGGVEHRLAECN 240
Db 586 FRKHPEATYSCGSGPWTTPCLVDYPRYLWHYPCTTNTYTFKIRMYVGGVEHRLAECN 645
QY 241 WTRGERCDLEDRDR 254
Db 646 WTRGERCDLEDRDR 659

RESULT 9
S19875
genome polyprotein - hepatitis C virus (isolate JK3) (fragment)
N;Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein
C;Species: hepatitis C virus
A;Variety: isolate JK3
C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 26-Aug-1999
C;Accession: S19875
R;Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
Submitted to the EMBL Data Library, September 1991
A;Description: Sequence analysis of putative structural regions of Hepatitis C Virus

A:Reference number: S18029

A:Accession: S19875

A:Molecule type: genomic RNA

A:Residues: 1-782 <HON>

A:Cross-references: EMBL:X61592; NID:G59482; PIDN:CAA43789.1; PID:G59483

A:Experimental source: isolate JK5

C:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology

C:Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural

F:1-191/Product: core protein status predicted <MAT1>

F:192-383/Product: core protein status predicted <MAT2>

F:384-733/Product: envelope protein 1 #status predicted <MAT3>

F:734-782/Product: NS1/E2 protein #status predicted <MAT4>

Query Match 83.3%; Score 1244; DB 2; Length 782;

Best Local Similarity 79.5%; Pred. No. 8.6e-94;

Matches 202; Conservative 25; Mismatches 27; Indels 0; Gaps 0;

QY 1 GAKQWQLINTNGSWHLNSTALNCNDSLNTGLAGLFYHHKFNSSGCPERLASCRPLTDF 60

DB 406 GPSONIQIQLVNSNGSWHLINRTALSCNDSLSKTGFLAALFYTHKFENASGCPERWASCRSDITF 465

QY 61 DOGNGPISYANGSGDQRPYCHWYPPKPGIVPAKSVCGPVYCFPTSPVVVGTDRSGAP 120

DB 466 DOGNGPITHVVPNTDQKPYCHWYAPRCGIVPASQVCGPVYCFPTSPVVVGTDRSGAP 525

QY 121 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCAGPVCVIGGAGNNTLHCPTDC 180

DB 526 TTYWGENETDVLNTRPPLGNWFGCTWMNSTGFTKVCAGPVCVIGGAGNNTLHCPTDC 585

QY 181 FRKHDPATYSRCGSGPWITPRCLVDYPRYLWHYPCITNTYFIKIRMYVGGVEHRLAECN 240

DB 586 FRKHPEATYTKCGSGPWLTPRCMVDPYRLWHYPCITNTYFIKIRMYVGGVEHRLAECN 645

QY 241 WTRGERCDLEDRDR 254

DB 646 WTRGERCDLSDRDR 659

RESULT 10

JC5620

genome polyprotein - hepatitis C virus (isolate EUH1480)

N:Contains: capsid protein C; envelope protein M; hepatitis C virus (fragment)

C:Species: hepatitis C virus

C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-May-2000

A:Accession: JC5620

R:Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.

Biochem. Biophys. Res. Commun. 236, 44-49, 1997

A:Title: The complete coding sequence of hepatitis C virus genotype 5a, the predominant

A:Reference number: JC5620; MUID:97366593

A:Accession: JC5620

A:Molecule type: mRNA

A:Residues: 1-3014 <CHA>

A:Cross-references: GB:Y13184

A:Experimental source: genotype 5a, which predominates in South Africa

A:Note: the translation of the nucleotide sequence is not complete in this paper

C:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology

C:Keywords: ATP; glycoprotein; hydrolase; P-loop; polyprotein; serine proteinase; trans

F:2-115/Product: capsid protein C #status predicted <CPC>

F:116-191/Product: envelope protein M #status predicted <EPM>

F:192-389/Product: major envelope protein E #status predicted <MEE>

F:384-408/Region: hypervariable #status predicted

F:390-730/Product: nonstructural protein NS1 #status predicted <NS1>

F:731-1007/Product: nonstructural protein NS2 #status predicted <NS2>

F:1008-1616/Product: hepatitis C virus genome polyprotein; serine proteinase; trans

F:1231-1494/Domain: DEAD/H box helicase homology <DEAD>

F:1231-1238/Region: nucleotide-binding motif A (P-loop)

F:1313-1318/Region: nucleotide-binding motif B

F:1317-1320/Region: DEH motif

F:1617-1863/Product: nonstructural protein NS4a #status predicted <N4A>

F:1864-2014/Product: nonstructural protein NS4b #status predicted <N4B>

F:2015-3014/Product: nonstructural protein NS5 #status predicted <NS5>

F:2210-2249/Region: interferon sensitivity determining #status predicted

Query Match 83.2%; Score 1242.5; DB 1; Length 3014;

Best Local Similarity 78.8%; Pred. No. 4.6e-93;

Matches 201; Conservative 26; Mismatches 27; Indels 1; Gaps 1;

QY 1 GAKQWQLINTNGSWHLNSTALNCNDSLNTGLAGLFYHHKFNSSGCPERLASCRPLTDF 60

DB 406 GPQQLQFVNTNGSWHLNSTALNCNDSLNTGLAGLFYHHKFNSSGCPERWASCRPLAEE 465

QY 61 DOGNGPISYANGSGDQRPYCHWYPPKPGIVPAKSVCGPVYCFPTSPVVVGTDRSGA 119

DB 466 DOGNGPISYATISGSDDKPYCHWYPPKPGIVPAKSVCGPVYCFPTSPVVVGTDRSG 525

QY 120 PTYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCAGPVCVIGGAGNNTLHCPT 179

DB 526 PTYWSGNETDILLNTRPAGNWFCTWMNSTGFTKVCAGPVCVIGGAGNNTLHCPT 585

QY 180 CFRKHDPATYSRCGSGPWITPRCLVDYPRYLWHYPCITNTYFIKIRMYVGGVEHRLAEC 239

DB 586 CFRKHDPATYSRCGSGPWLTPRCVLHYPRYLWHYPCITNTYFIKIRMYVGGVEHRLAEC 645

QY 240 NWTGERCDLEDRDR 254

DB 646 NWTGERCDLEDRDR 660

RESULT 11

S19876

genome polyprotein - hepatitis C virus (isolate JK5) (fragment)

N:Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein

C:Species: hepatitis C virus

A:Variety: isolate JK5

C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 26-Aug-1999

C:Accession: S19876

R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.

submitted to the EMBL Data Library, September 1991

A:Description: Sequence analysis of putative structural regions of Hepatitis C Virus

A:Reference number: S18029

A:Accession: S19876

A:Molecule type: genomic RNA

A:Residues: 1-782 <HON>

A:Cross-references: EMBL:X61595; NID:G59486; PIDN:CAA43792.1; PID:G59487

A:Experimental source: isolate JK5

C:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology

C:Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructur

F:1-191/Product: core protein status predicted <MAT1>

F:192-383/Product: envelope protein 1 #status predicted <MAT2>

F:384-733/Product: NS1/E2 protein #status predicted <MAT3>

F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match

82.7%; Score 1236; DB 2; Length 782;

Best Local Similarity 79.5%; Pred. No. 3.9e-93;

Matches 202; Conservative 25; Mismatches 27; Indels 0; Gaps 0;

QY 1 GAKQWQLINTNGSWHLNSTALNCNDSLNTGLAGLFYHHKFNSSGCPERLASCRPLTDF 60

DB 406 GPSONIQIQLVNSNGSWHLINRTALSCNDSLSKTGFLAALFYTHKFENASGCPERWASCRSDITF 465

QY 61 DOGNGPISYANGSGDQRPYCHWYPPKPGIVPAKSVCGPVYCFPTSPVVVGTDRSGAP 120

DB 466 DOGNGPITHVVPNTDQKPYCHWYAPRCGIVPASQVCGPVYCFPTSPVVVGTDRSGAP 525

QY 121 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCAGPVCVIGGAGNNTLHCPTDC 180

DB 526 TYNWGANETDVLNTRPPOGNWFGCTWMNSTGFTKVCAGPVCVIGGAGNNTLHCPTDC 585

QY 181 FRKHDPATYSRCGSGPWITPRCLVDYPRYLWHYPCITNTYFIKIRMYVGGVEHRLAECN 240

DB 586 FRKHPEATYTKCGSGPWLTPRCMVDPYRLWHYPCITNTYFIKIRMYVGGVEHRLAECN 645

C;Accession: S18032
R;Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
submitted to the EMBL Data Library, September 1991
A;Description: Sequence analysis of putative structural regions of Hepatitis C Virus iso
A;Reference number: S18029
A;Accession: S18032
A;Molecule type: genomic RNA
A;Residues: 1-782 <HON>
A;Cross-references: EMBL: X61594
A;Experimental source: isolate JK4
C;Superfamily: hepatitis C virus genome polypeptide; DEAD/H box helicase homology
C;Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural
F;1-191/Product: core protein #status predicted <NAT1>
F;192-383/Product: envelope protein 1 #status predicted <NAT2>
F;384-733/Product: NS1/E2 protein #status predicted <NAT3>
F;734-782/Product: nonstructural protein 2 (fragment) #status predicted <NAT4>

Query Match	82.2%	Score 1228;	DB 2;	Length 782;
Best Local Similarity	79.1%	Pred. No. 1.7e-92;		
Matches 201; Conservative	23;	Mismatches 30;	Indels 0;	Gaps 0;

[illegible]

RESULT 15

S18031 genome polyprotein - hepatitis C virus (isolate JK2) (fragment)
 N:Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein
 C:Species: hepatitis C virus
 A:Variety: isolate JK2
 C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 29-May-1998
 C:Accession: S18031
 R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
 Submitted to the EMBL Data Library, September 1991
 A:Description: Sequence analysis of putative structural regions of Hepatitis C Virus iso
 A:Reference number: S18029
 A:Accession: S18031
 A:Molecule type: genomic RNA
 A:Residues: 1-782 <HON>
 A:Cross-references: EMBL:X61593
 A:Experimental source: isolate JK2
 S:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology
 C:Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural
 I:1-191/Product: core protein status predicted <MAT1>
 I:192-383/Product: envelope protein 1 #status predicted <MAT2>
 I:384-733/Product: NS1/E2 protein #status predicted <MAT3>
 I:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 80.3%; Score 1199; DB 2; Length 782;
Best Local Similarity 76.8%; Pred. No. 4e-90;
Matches 195; Conservative 27; Mismatches 32; Indels 0; Gaps 0;

QY 1 GAKQNVQLINTNGSWHLNSTALNCNDSLNTGWLGLFYHHKFNSSGGCPERLASCRPLTDF 60

Db	406	GPSKIQIIVNSGRHINRTALSCNLSKLTGFUAALFYTHKFNAAQCPERMASCCSIDTF	465
Qy	61	DQGNPTISYANGSGPOQRPCYCHWYPPKPCIVPAKSVCGPVGFTSPVVVGTTRDSGAP	120
Db	466	DQGNPTITHVVPNSTDOKPYCWHYAPRPCIVPASVCGPVGFTSPVVVGTTRDEGAP	525
Qy	121	TYSGENDTDFVLNNTRPPLGNHNFCTWMNSTGFTKVCACAPCVLGGAGNNTLHCPTDC	180
Db	526	TYNKGANETDVLNNTRPPQGNHNFCTWMNSTGFTKTCGGPPCNTGGVGNNTLTCPTDC	585
Qy	181	FRKHDPATYSCSGSGPWITPRCLVDYPYRLWHVPCNTINVTIFKIRMYVGVGEHRLTAACN	240
Db	586	FRKHPEATYKCGSGGPMWLPTRCMVHPYPRLWHVPCNTVNFVFKVRMYVGVGEHRLTAACN	645
Qy	241	WTRGERCDLEDRDR	254
Db	646	WTRGERCNLEDRDR	659

Search completed: March 6, 2001, 11:53:21
Job time: 158 sec

us-09-407-430-3.rpr

Tue Mar 6 12:02:35 2001

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 11:53:43 ; Search time 25.34 Seconds
(without alignments)
323.706 Million cell updates/sec

Title: US-09-407-430-3
Perfect score: 1494
Sequence: 1 GAKQNVQLINTSGWHLST.....LEACNWTGRGCDLEDNRDR 254

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.4

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1494	100.0	3011	1	POLG_HCV1
2	1398	93.6	3011	1	POLG_HCV1
3	1273	85.2	3010	1	POLG_HCV1
4	1254	83.9	3010	1	POLG_HCV1
5	1253	83.9	3010	1	POLG_HCV1
6	1234	82.6	3010	1	POLG_HCV1
7	1170	78.3	3033	1	POLG_HCV1
8	1158	77.5	737	1	POLG_HCV1
9	1156	77.4	3033	1	POLG_HCV1
10	1147	76.8	737	1	POLG_HCV1
11	511	34.2	520	1	POLG_HCV1
12	511	34.2	520	1	POLG_HCV1
13	475	31.8	513	1	POLG_HCV1
14	173	11.6	321	1	POLG_HCV1
15	171	11.4	138	1	POLG_HCV1
16	167	11.2	321	1	POLG_HCV1
17	153	10.2	309	1	POLG_HCV1
18	110.5	7.4	1700	1	BAR3_CHITE
19	107	7.2	1046	1	PSTA_DICDI
20	103	6.9	1172	1	TSP2_DICDI
21	97.5	6.5	640	1	UROM_HUMAN
22	96.5	6.5	752	1	CO2_HUMAN
23	94	6.3	555	1	DP87_DICDI
24	93.5	6.3	1172	1	TSP2_HUMAN
25	91.5	6.1	571	1	HEWA_P12H
26	91.5	6.1	571	1	HEWA_P12H
27	91	6.1	2871	1	FBNI_HUMAN
28	90	6.0	1170	1	TSP2_BOVIN
29	90	6.0	2871	1	FBNI_BOVIN
30	88.5	5.9	784	1	YAV2_XANCV
31	88.5	5.9	1077	1	SM5A_MOUSE
32	87	5.8	1178	1	TSP1_CHICK
33	87	5.8	2871	1	FBNI_BOVIN

RESULT 1

POLG_HCV1

AC

P26664;

DT

01-AUG-1992

(Rel. 23, Created)

DE

GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN) (P22); ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35); ENVELOPE GLYCOPROTEIN E2 (GP68) (GP70) (NS1); PROTEIN P7; NONSTRUCTURAL PROTEIN NS2 (P21) (EC 3.4.22.-); PROTEIN P7; NONSTRUCTURAL PROTEIN NS3 (P70) (EC 3.4.21.-); NONSTRUCTURAL PROTEIN NS4A (P4); NONSTRUCTURAL PROTEIN NS4B (P27); NONSTRUCTURAL PROTEIN NS5A (P56); NONSTRUCTURAL PROTEIN NS5B (P66) (P70) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48)].

DE

Hepatitis C virus (isolate 1) (HCV).

OS

Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.

OC

Hepacivirus.

RN

[1]

SEQUENCE FROM N.A.

MEDLINE-91172826; PubMed-1848704;

Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C., Gallegos C., Coit D., Medina-Selby A., Barr P.J., Welner A.J., Bradley D.W., Kuo G., Houghton M.;

"Genetic organization and diversity of the hepatitis C virus."

Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).

CC

FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.

CC

NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.

CC

SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.

CC

SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

CC

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

CC

EMBL; M62321; AAA45676.1;

PIR; A39166; GNWVC3.

DR

HSSP; P27958; 1HEI.

DR

MEROPS; S29.001;

DR

MEROPS; U39.001;

DR

INTERPRO; IPR000745;

DR

INTERPRO; IPR001490;

DR

INTERPRO; IPR002166;

DR

INTERPRO; IPR002518;

DR

INTERPRO; IPR002519;

DR

INTERPRO; IPR002521;

DR

INTERPRO; IPR002522;

DR

INTERPRO; IPR002531;

DR

INTERPRO; IPR002868;

Q24400 drosophila
P05160 homo sapien
P31747 bacillus sp
Q27969 bos taurus
P48733 bos taurus
Q39692 daucus caro
P12625 alcaligenes
P56680 lumicola in
Q61483 us musculus
P75750 scherichia
P07996 homo sapien
P22549 dictyostell

34 86.5 5.8 495 1 MLP2_DROME
35 85.5 5.7 661 1 F13B_HUMAN
36 85.5 5.7 718 1 CDGT_BACSS
37 84.5 5.7 532 1 AD50_BOVIN
38 84.5 5.7 543 1 UROM_BOVIN
39 82.5 5.5 592 1 INV2_DAUCA
40 82 5.5 488 1 PHB_ALCFA
41 81.5 5.5 402 1 GUN1_HUMIN
42 81.5 5.5 722 1 DL11_MOUSE
43 81.5 5.5 818 1 YBGO_ECOLI
44 81.5 5.5 1170 1 TSP1_HUMAN
45 81 5.4 237 1 IPDE_DICDI

ALIGNMENTS

STANDARD; PRT; 3011 AA.

DR PFAM; PF01560; HCV_NS1; 1.
 DR PFAM; PF01538; HCV_NS2; 1.
 DR PFAM; PF01006; HCV_NS4a; 1.
 DR PFAM; PF01001; HCV_NS4b; 1.
 DR PFAM; PF01506; HCV_NS5a; 1.
 DR PFAM; PF00998; HCV_RDRP; 1.
 DR PFAM; PF01543; HCV_capsid; 1.
 DR PFAM; PF01542; HCV_core; 1.
 DR PFAM; PF01539; HCV_env; 1.
 KW Polyprotein; glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural
 FT INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3011
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 407 407
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 476 476
 FT CARBOHYD 532 532
 FT CARBOHYD 540 540
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2364 2364
 FT CARBOHYD 2789 2789
 FT SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;
 Query Match 100.08; Score 1494; DB 1; Length 3011;
 Best Local Similarity 100.08; Pred. No. 7.8e-121;
 Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GAKONVQLINTGSHLNTALNCNDLSLNTGLAGLFFYHHKFNSSGCPERLASCRPLTDF 60
 DB 406 GAKONVQLINTGSHLNTALNCNDLSLNTGLAGLFFYHHKFNSSGCPERLASCRPLTDF 465
 QY 61 DQGWGPISYANGSGDPQRYCHYPPKPCGIVPAKSVCGPVYCFPTSPVVGVTTDRSGAP 120
 DB 466 DQGWGPISYANGSGDPQRYCHYPPKPCGIVPAKSVCGPVYCFPTSPVVGVTTDRSGAP 525
 QY 121 TYSWGENDDTVFLNTRPLGNWFCTWNSTGTFKVCAGPCVITGGAGNNTLHCPTDC 180
 DB 526 TYSWGENDDTVFLNTRPLGNWFCTWNSTGTFKVCAGPCVITGGAGNNTLHCPTDC 585
 QY 181 FRKHPDATSRCSGGPWITPRCLVDYPIRLWHYPCTINTYIFKIRMYGVGVHREAAACN 240
 DB 586 FRKHPDATSRCSGGPWITPRCLVDYPIRLWHYPCTINTYIFKIRMYGVGVHREAAACN 645
 QY 241 WTRGERCDLEDRDR 254

DB 646 WTRGERCDLEDRDR 659

RESULT 2

POLG_HCVH STANDARD; PRT; 3011 AA.
 ID POLG_HCVH
 AC P27958;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 38, Last annotation update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN) (P22);
 DE ENVELOPE GLYCOPROTEIN E1 (GP32); ENVELOPE GLYCOPROTEIN E2
 DE (GP68) (GP70) (NS1); PROTEIN P7; NONSTRUCTURAL PROTEIN NS2 (P21)
 DE (EC 3.4.22.-); PROTEINASE/HELICASE NS3 (P70) (EC 3.4.21.-);
 DE NONSTRUCTURAL PROTEIN NS4A (P4); NONSTRUCTURAL PROTEIN NS4B (P27);
 DE NONSTRUCTURAL PROTEIN NS5A (P56); NONSTRUCTURAL PROTEIN NS5B (P66)
 DE (P70) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate H) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 RN Hepacivirus.
 CC [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-92052256; PubMed-1658800;
 RA Inchauspe G., Zebade S., Lee D.H.H., Sugitani M., Nasoff M.,
 RA Prince A.M.;
 RA "Genomic structure of the human prototype strain H of hepatitis C
 RT virus: comparison with American and Japanese isolates.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
 RN [2]
 RX X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
 RP MEDLINE-97331322; PubMed-9187654;
 RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
 RA "Structure of the hepatitis C virus RNA helicase domain.";
 RL Nat. Struct. Biol. 4:463-467(1997).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
 RX MEDLINE-98134321; PubMed-9493270;
 RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
 RA Murcko M.A., Lin C., Caron P.R.;
 RA "Hepatitis C virus NS3 RNA helicase domain with a bound
 RT oligonucleotide: the crystal structure provides insights into the mode
 of unwinding.";
 RL Structure 6:89-100(1998).
 CC -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
 CC -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
 CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
 CC -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
 CC ACTIVATION OF NS3.
 CC -!- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
 CC -!- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
 CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
 CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
 CC -!- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
 CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
 CC -!- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY C18.
 CC -!- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M67463; AAA45534.1;
 DR PIR; A36814; GNMVCH.
 DR PDB; 1HEI; 25-NOV-98.
 DR PDB; 1A1V; 16-FEB-99.
 DR MEROPS; S29.001; -.

D	B		526	TYSWANDTDVFLVNTRPLGNWFCTWMNSTGFTKVCAPPCVIGGVGNNTLLCPTDC	583
O	Y		181	FRKHDPATYSRCGSGPWITPRCLVDYPYRLWHYPCTINVTIFKIRMYGGVEHRLAAACN	240
D	B		586	FRKYPEATYSRCGSGPRIIPRCMWDYPYRLWHYPCTINVTIFKIRMYGGVEHRLAAACN	645
O	Y		241	WTRGERCDLEDNRD	254
D	B		646	WTRGERCDLEDNRD	659

RESULT 3
POLG_HCVTW STANDARD: PRT: 3010 AA.

```

AC      P29846;
AD      01-APR-1993 (Rel. 25, Created)
DT      01-APR-1993 (Rel. 25, Last sequence update)
DD      15-JUL-1999 (Rel. 38, Last annotation update)
DE      GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN) (P22);
DE      ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35); ENVELOPE GLYCOPROTEIN E2
DE      (GP68) (GP70) (NS1); PROTEIN P7; NONSTRUCTURAL PROTEIN NS2 (P21)
DE      (EC 3.4.22.-); PROTEASE/HELICASE NS3 (P70) (EC 3.4.21.-);
DE      NONSTRUCTURAL PROTEIN NS4A (P4); NONSTRUCTURAL PROTEIN NS4B (P27);
DE      NONSTRUCTURAL PROTEIN NS5A (P56); NONSTRUCTURAL PROTEIN NS5B (P66)
DE      (P70) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48)].
OS      Hepatitis C virus (Isolate Taiwan) (HCV).
OC      Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC      Hepacivirus.
RN      [1]
RR      SEQUENCE FROM N.A.
RX      MEDLINE=92230206; PubMed=1314449;
RT      Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;
RT      "The Taiwanese hepatitis C virus genome: sequence determination and
RT      mapping the 5'-termini of viral genomic and antigenomic RNA.";
RT      Virology 188:102-113(1992).
CC      -|- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC      HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC      NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC      -|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY
CC      LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC      PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC      PROTEIN C AND MRNA.
CC      -|- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collabora-
CC      tion between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      at the European Bioinformatics Institute. There are no restrictions on
CC      use by non-profit institutions as long as its content is in no
CC      way modified and this statement is not removed. Usage by and for commer-
CC      cial entities requires a license agreement (see http://www.isb-sib.ch/announ-
CC      ce or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: M84754; -, NOT_ANNOTATED_CDS.
CC      PIR: A40244; GNVVTV.
CC      HSP: P27958; 1HEI.
CC      MEROPS: S29.001; -.
CC      MEROPS: U39.001; -.
CC      INTERPRO: IPR000745; -.
CC      INTERPRO: IPR001490; -.
CC      INTERPRO: IPR002166; -.
CC      INTERPRO: IPR002518; -.
CC      INTERPRO: IPR002519; -.
CC      INTERPRO: IPR002521; -.
CC      INTERPRO: IPR002522; -.
CC      INTERPRO: IPR002531; -.
CC      INTERPRO: IPR002868; -.
CC      PFAM: PF01560; HCV_NS1; 1.
CC      PFAM: PF01538; HCV_NS2; 1.
CC      PFAM: PF01006; HCV_NS4a; 1.
CC      PFAM: PF01001; HCV_NS4b; 1.
CC      PFAM: PF01506; HCV_NS5a; 1.
CC      PFAM: PF00998; HCV_RDRP; 1.
CC      PFAM: PF01543; HCV_GNSS4; 1.

```

	Query Match	93.6%	Score 1398;	DB 1;	Length 3011;
	Best Local Similarity	93.3%	Pred. No. 1.5e-112;		
	Matches 237;	Conservative	9;	Mismatches 8;	Indels 0; Gaps 0;
QY	1	CAKQNVQLINTNGSWHLNSTALNCNDSLNTGWLGLFYHHKFNSSGCPRLASCRPLTDF	60		
dbb					
dbb	406	CAKQNIQLINTNGSWHLNSTALNCNDSLNTGWLGLFYHHKFNSSGCPRLASCRRLTDF	465		
QY	61	DQGWCPISYANGSPDORPCWHPKPCICVPAKSVCGPVYCFPTSPVVVGTDRSGAP	120		
dbb	466	AQGMPISYANGSLDERPCWHPKPCICVPAKSVCGPVYCFPTSPVVVGTDRSGAP	525		
QY	121	TVSGENDTDVEVLNNTRPPLGNWFEGCTWNNSTGTFKVCAGPPCVIGGAGNNLTLCPTDC	180		
dbb					

DR PFAM: PF01542; HCV_core; 1.
DR PFAM: PF01539; HCV_env; 1.
DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Glycoprotein; Helicase; ATP-binding;
KW Core protein; Coat protein; Envelope protein; Hydrolase; Serine protease.
KW Intramembrane; Nonstructural
FT INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 729
FT CHAIN 730 1006
FT CHAIN 1007 1615
FT CHAIN 1616 1862
FT CHAIN 1863 2013
FT CHAIN 2014 3010
FT CHAIN TRANSMEM 347 369
FT ACT_SITE 1083 1083
FT ACT_SITE 1107 1107
FT ACT_SITE 1165 1165
FT NP_BIND 1230 1237
FT SITE 1316 1319
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 233 233
FT CARBOHYD 234 234
FT CARBOHYD 250 250
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 532 532
FT CARBOHYD 540 540
FT CARBOHYD 556 556
FT CARBOHYD 576 576
FT CARBOHYD 623 623
FT CARBOHYD 645 645
FT CARBOHYD 2041 2041
FT CARBOHYD 2077 2077
FT CARBOHYD 2240 2240
FT CARBOHYD 2529 2529
FT CARBOHYD 2788 2788
SQ SEQUENCE 3010 AA; 327047 MW; AAD267D55CDFE215 CRC64;
Query Match 85.2%; Score 1273; DB 1; Length 3010;
Best Local Similarity 81.9%; Pred No. 8, 7e-102;
Matches 208; Conservative 19; Mismatches 27; Indels 0; Gaps 0;
Qy 1 GAKONVQLINTGSHWLNSTALNCNDSLNTGWLGLFYHHKNSGCPERLASCRPLTDF 60
Db 406 GASQKIQLINTGSHWLNSTALNCNDSLNTGWLGLFYHHKNSGCPERLASCRPLTDF 60
Qy 61 DQGMPTISYANGSDPDRPCWHYPPKPGICVPAKSCVPGVCFTPSPVVVGTITDRSGAP 120
Db 466 DQGMPTITTEADTQDQPCWHYAPRCGIVPASPQCGPVCFCTPSPVVVGTITDRSGAP 525
Qy 121 TYSGENDTDFVLNNTRPPLGNWFGCTWMNSTGFTKVGAPPVCVIGGAGNNTLHCPTDC 180
Db 526 TYSGENETDVLINNTRPQGNWFGCTWMNSTGFTKTCGGPPCNIGGGNNTLVCPTDC 585
Qy 181 FRKHDPATYSCGGSPWITPRCLVDYPRLWHVPCNTIFKIRMYGVGVEHRLAECN 240
Db 586 FRKHDPATYTCGGSPWITPRCLVDYPRLWHVPCNTIFKIRMYGVGVEHRLAECN 645
Qy 241 WTRGERCDLEDRDR 254
Db 646 WTRGERCDLEDRDR 659
RESULT 4
POLG_HCVBK

Query Match	83.9%;	Score 1253;	DB 1;	Length 3010;
Best Local Similarity	79.5%;	Pred. No. 4.6e-100;		
Matches 202: Conservative	25;	Mismatches 27;	Indels 0;	Gaps 0;

Qy	1	GAKONTVOLINTNSGWHLNSTALNCNDSLNTGWLAGLFYHHKSNSSCCPRLASCRPLTDF	60
Db	406	GPSOKITQVNTNSWHINTALNCNDSLOTFQTAALFYAHRFNASGCPERMASCRPIDF	465
Qy	61	DQWGPISYANGSGPDORPYCWHYPPKCGIVPAKSVCGPVYCFTPSPVVVGTDRSGAP	120
Db	466	AQWGMPITHDMPSSDQRPYCYWHIYAPRCGIVPASQVCGPVYCFTPSPVVVGTDRFGAP	525
Qy	121	TYSWGENDTDFVLYNTRPPLGNWFGCTWMNSTGFKVCGAPPVCTGGAGNNTLHCPTDC	180
Db	526	TYSGENETDVLNSTRPPOGNWFGCTWMNSTGFKTCGGPPCNLGGVGNNTLVCPTDC	585
Qy	181	FRKHDPATYSCSGSGPWITPRCLVDPPYRLWHYPCTINTYIFKIRMYVGVGVEHRLAACN	240
Db	586	FRKHEATYTKCSGSGPWLTPRCMVDYPPYRLWHYPECTVNETVFKVRMYVGVGVEHRLAACN	645
Qy	241	WTRGERCDLEDORR	254

646 WTRGCRCDLEDRDR 659

RESULT 6
POLG.HCVJT
ID POLG.HCVJT STANDARD; PRT; 3010 AA.
AC Q00269;
AD 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN) (P22):
DE ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35): ENVELOPE GLYCOPROTEIN E2
DE (GP68) (GP70) (NS1): PROTEIN P7: NONSTRUCTURAL PROTEIN NS2 (P21)
DE (EC 3.4.22.-); PROTEASE/HELICASE NS3 (P70) (EC 3.4.21.-);
DE NONSTRUCTURAL PROTEIN NS4A (P4); NONSTRUCTURAL PROTEIN NS4B (P27):
DE NONSTRUCTURAL PROTEIN NS5A (P56); NONSTRUCTURAL PROTEIN NS5B (P66)
DE (P70) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-JT) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae.
OC Hepacivirus.
OC [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=92295714; PubMed=1318627;
RX Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,
RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
RT "Molecular cloning of hepatitis C virus genome from a single Japanese
RT carrier: sequence variation within the same individual and among
RT infected individuals.";
RL Virus Res. 23:39-53(1992).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collabora-
CC between the Swiss Institute of Bioinformatics and the EMBL Outsta-
CC the European Bioinformatics Institute. There are no restrictions c-
CC use by non-profit institutions as long as its content is in n-
CC modified and this statement is not removed. Usage by and for comm-
CC entities requires a license agreement (see <http://www.isb-sib.ch/annou>-
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; J11168; BAA01943.1; -
CC PIR; A45573; A45573.
CC DR HSP; P27958; 1HEI.
CC DR MEROPS; S29.001; -.
CC DR MEROPS; U39.001; -.
CC DR INTERPRO; IPR000745; -
CC DR INTERPRO; IPR001490; -
CC DR INTERPRO; IPR002166; -
CC DR INTERPRO; IPR002518; -
CC DR INTERPRO; IPR002519; -
CC DR INTERPRO; IPR002521; -
CC DR INTERPRO; IPR002522; -
CC DR INTERPRO; IPR002531; -
CC DR INTERPRO; IPR002868; -
CC DR PFAM; PF01560; HCV_NS1; 1.
CC DR PFAM; PF01538; HCV_NS2; 1.
CC DR PFAM; PF01006; HCV_NS4a; 1.
CC DR PFAM; PF01001; HCV_NS4b; 1.
CC DR PFAM; PF00998; HCV_RGRP; 1.
CC DR PFAM; PF01543; HCV_capsid; 1.
CC DR PFAM; PF01542; HCV_core; 1.
CC DR PFAM; PF01539; HCV_env; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
FT INIT_MET 1 1
FT REMOVED FROM CAPSID PROTEIN C BY THE

FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 729
 FT CHAIN 730 1006
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT CHAIN 3010 326573
 FT TRANSMEM 347 369
 FT ACT_SITE 1083 1083
 FT ACT_SITE 1107 1107
 FT ACT_SITE 1165 1165
 FT NP_BIND 1230 1237
 FT SITE 1316 1319
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 250 250
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 532 532
 FT CARBOHYD 540 540
 FT CARBOHYD 556 556
 FT CARBOHYD 576 576
 FT CARBOHYD 623 623
 FT CARBOHYD 645 645
 FT CARBOHYD 2041 2041
 FT CARBOHYD 2077 2077
 FT CARBOHYD 2240 2240
 FT CARBOHYD 2529 2529
 FT CARBOHYD 2788 2788
 SQ SEQUENCE 3010 AA; 326573 MW; 94A1C774350642BB CRC64;

Query Match 82.6%; Score 1234; DB 1; Length 3010;
 Best Local Similarity 79.9%; Pred. No. 2e-98;
 Matches 203; Conservative 19; Mismatches 32; Indels 0; Gaps 0;

QY 1 GAKQNYQLINTNGSHLNTALNCDSLNTGWLGLFYHHKFNSSCCPRLASCRPLTDF 60
 DB 406 GPAQRLQINTNGSHLNTALNCDSLNTGWLGLFYHHKFNSSCCPRLASCRPLTDF 60
 QY 61 DQGWGPIYANGSGDQRPYCHWHPKPGIYPAKSVGPGVYCFPTSPVVGCTDRSGAP 120
 DB 466 AOCWGPITVTEPRDLDRPYCHWHPKPGIYPAKSVGPGVYCFPTSPVVGCTDRSGAP 525
 QY 121 TYSWGENDTDFVLLNTRPPLGNWFGCTWNSGTFTKVGCGAPCVIGGAGNNLHCPTDC 180
 DB 526 TYNWGANETDVLNTRPPLGNWFGCTWNSGTFTKVGCGAPCVIGGAGNNLHCPTDC 585
 QY 181 FRKHDPATYSRSGSGPWITPRLVDYPRYLWHPCTINTYIKIRMYGCVGVEHLEAACN 240
 DB 586 FRKHDPATYSRSGSGPWITPRLVDYPRYLWHPCTINTYIKIRMYGCVGVEHLEAACN 645
 QY 241 WTRGERCDLEDRDR 254
 DB 646 WTRGERCDLEDRDR 659

RESULT 7
 POLG_HCVJ8
 ID POLG_HCVJ8 STANDARD; PRT; 3033 AA.
 AC P26661;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE GENOME POLYPROTEIN (CONTAINS: CAPSID PROTEIN C (CORE PROTEIN) (P22);
 ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35); ENVELOPE GLYCOPROTEIN E2

DE DE (GP68) (GP70) (NS1); PROTEIN P7; NONSTRUCTURAL PROTEIN NS2 (P21)
 DE (BC 3.4.22.-); PROTEASE/HELICASE NS3 (P70) (EC 3.4.21.-);
 DE NONSTRUCTURAL PROTEIN NS4A (P4); NONSTRUCTURAL PROTEIN NS4B (P27);
 DE NONSTRUCTURAL PROTEIN NS5A (P56); NONSTRUCTURAL PROTEIN NS5B (P66)
 OS (P70) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48)].
 OC Hepatitis C virus (isolate HC-J8) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-92230232; PubMed-1314459;
 RA Okamoto H., Kurai K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T.,
 RA Fukuda S., Tsuda F., Mishiro S.:
 RT "Full-length sequence of a hepatitis C virus genome having poor
 RT homology to reported isolates: comparative study of four distinct
 RT genotypes";
 RL Virology 188:331-341(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@sib-sib.ch).
 DR EMBL: D10988; BAA01761.1; -
 DR PIR: A40250; GNMVJ8.
 DR HSP: P27958; 1HEI.
 DR MEROPS: S29.001; -
 DR INTERPRO: IPR000745; -
 DR INTERPRO: IPR001490; -
 DR INTERPRO: IPR002166; -
 DR INTERPRO: IPR002518; -
 DR INTERPRO: IPR002519; -
 DR INTERPRO: IPR002521; -
 DR INTERPRO: IPR002522; -
 DR INTERPRO: IPR002531; -
 DR INTERPRO: IPR002868; -
 DR PFAM: PF01560; HCV_NS1; 1.
 DR PFAM: PF01538; HCV_NS2; 1.
 DR PFAM: PF01006; HCV_NS4a; 1.
 DR PFAM: PF01001; HCV_NS4b; 1.
 DR PFAM: PF01506; HCV_NS5a; 1.
 DR PFAM: PF00998; HCV_RdRp; 1.
 DR PFAM: PF01543; HCV_Capsid; 1.
 DR PFAM: PF01542; HCV_Core; 1.
 DR PFAM: PF01539; HCV_env; 1.
 DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 DR Core protein; Coat protein; Envelope protein; Helicase; ATP-binding
 DR Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
 FT INIT_MET 1 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT CHAIN 347 369
 FT TRANSMEM 347 369
 FT ACT_SITE 1087 1087

Tue Mar 6 12:02:35 2001

FT ACT_SITE 1111 1111 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT_SITE 1169 1169 ATP (POTENTIAL).

FT NP_BIND 1234 1241 DECH BOX.

FT SITE 1320 1323

FT CARBOHYD 196 196 N-LINKED (GLCNAC. . . .) (POTENTIAL).

FT CARBOHYD 209 209 N-LINKED (GLCNAC. . . .) (POTENTIAL).

FT CARBOHYD 233 233 N-LINKED (GLCNAC. . . .) (POTENTIAL).

FT CARBOHYD 299 299 N-LINKED (GLCNAC. . . .) (POTENTIAL).

FT CARBOHYD 305 305 N-LINKED (GLCNAC. . . .) (POTENTIAL).

FT CARBOHYD 417 417 N-LINKED (GLCNAC. . . .) (POTENTIAL).

FT CARBOHYD 423 423 N-LINKED (GLCNAC. . . .) (POTENTIAL).

FT CARBOHYD 430 430 N-LINKED (GLCNAC. . . .) (POTENTIAL).

FT CARBOHYD 448 448 N-LINKED (GLCNAC. . . .) (POTENTIAL).

FT CARBOHYD 477 477 N-LINKED (GLCNAC. . . .) (POTENTIAL).

FT CARBOHYD 534 534 N-LINKED (GLCNAC. . . .) (POTENTIAL).

FT CARBOHYD 542 542 N-LINKED (GLCNAC. . . .) (POTENTIAL).

FT CARBOHYD 558 558 N-LINKED (GLCNAC. . . .) (POTENTIAL).

FT CARBOHYD 578 578 N-LINKED (GLCNAC. . . .) (POTENTIAL).

FT CARBOHYD 627 627 N-LINKED (GLCNAC. . . .) (POTENTIAL).

FT CARBOHYD 649 649 N-LINKED (GLCNAC. . . .) (POTENTIAL).

FT CARBOHYD 1091 1091 N-LINKED (GLCNAC. . . .) (POTENTIAL).

FT CARBOHYD 2038 2038 N-LINKED (GLCNAC. . . .) (POTENTIAL).

FT CARBOHYD 2359 2359 N-LINKED (GLCNAC. . . .) (POTENTIAL).

FT CARBOHYD 2811 2811 N-LINKED (GLCNAC. . . .) (POTENTIAL).

FT CARBOHYD 3033 3033 AA; 330177 MW; 1A173E7E3381FD1A CRC64;

SQ SEQUENCE 3033 AA; 78.3%; Score 1170; DB 1; Length 3033;

Query Match 78.3%; Score 1170; DB 1; Length 3033;

Best Local Similarity 76.7%; Pred. No. 6.6e-93;

Matches 198; Conservative 22; Mismatches 34; Indels 4; Gaps 2;

QY 1 GAKQNVQLINTNGSHLNTALNCNDLSLNTGLAGLFYHHKFNSSGCGPERLASCRPLTDF 60

DB 406 GAKQNVQLINTNGSHLNTALNCNDLSLNTGLAGLFYHHKFNSSGCGPERLASCRPLTDF 465

QY 61 DOGWGPISYANG--SGPDORPCYCHYPPKPGCVIPAKSVGCVYCFPTSPVVVGTTRSG 118

DB 466 RIGWGTLEYETNVTNEEDMRPYCHYPPKPGCVIPAKSVGCVYCFPTSPVVVGTTRSG 525

QY 119 APTYSWGENDTDFVLNNTTRPPLGNFNGCTWMNSTGFTKVCAGPCVIGGAGNNTLH--C 176

DB 586 PTDCFRKHPDATYLRKAGGFWLTPCLVDYPIRLMHPCTVNTIFKRMVYGVGVEHRL 645

QY 237 AACNWTGRGCDLEDRDR 254

DB 646 AACNFTRGDCRLEDRDR 663

RESULT 8

POLG_HCVJ7 STANDARD; PRT; 737 AA.

ID POLG_HCVJ7

AC P27961

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN); MATRIX

DE PROTEIN (ENVELOPE PROTEIN M); MAJOR ENVELOPE PROTEIN E; NONSTRUCTURAL

DE PROTEINS NS1 AND NS2] (FRAGMENT).

OS Hepatitis C virus (isolate HC-37) (HCV).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=92230232; PubMed=1314459;

RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,

RA Tanaka T., Fukuda S., Tsuda F., Mishihiro S.,

RT "Full-length sequence of a hepatitis C virus genome having poor

RT homology to reported isolates: comparative study of four distinct

RT genotypes.";

RL VIROLOGY 188:331-341(1992).

CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE

CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.

CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.

CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A

CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:

CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

CC PROTEIN C AND MRNA.

CC -----

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; D10077; BAA00971.1; -

CC INTERPRO; IPR002519; -

CC INTERPRO; IPR002521; -

CC INTERPRO; IPR002522; -

CC INTERPRO; IPR002531; -

CC PFAM; PF01560; HCV_NSL; 1.

CC PFAM; PF01543; HCV_capsid; 1.

CC PFAM; PF01542; HCV_core; 1.

CC PFAM; PF01539; HCV_env; 1.

CC Polyprotein; Glycoprotein; Coat protein; Envelope protein;

CC Transmembrane; Nonstructural protein.

CC INIT_MET 1 1

CC CHAIN 1 115

CC CHAIN 116 191

CC CHAIN 192 383

CC CHAIN 384 733

CC CHAIN 734 >737

CC TRANSMEM 347 369

CC CARBOHYD 196 196 N-LINKED (GLCNAC. . . .) (POTENTIAL).

CC CARBOHYD 209 209 N-LINKED (GLCNAC. . . .) (POTENTIAL).

CC CARBOHYD 233 233 N-LINKED (GLCNAC. . . .) (POTENTIAL).

CC CARBOHYD 299 299 N-LINKED (GLCNAC. . . .) (POTENTIAL).

CC CARBOHYD 305 305 N-LINKED (GLCNAC. . . .) (POTENTIAL).

CC CARBOHYD 417 417 N-LINKED (GLCNAC. . . .) (POTENTIAL).

CC CARBOHYD 423 423 N-LINKED (GLCNAC. . . .) (POTENTIAL).

CC CARBOHYD 430 430 N-LINKED (GLCNAC. . . .) (POTENTIAL).

CC CARBOHYD 448 448 N-LINKED (GLCNAC. . . .) (POTENTIAL).

CC CARBOHYD 477 477 N-LINKED (GLCNAC. . . .) (POTENTIAL).

CC CARBOHYD 534 534 N-LINKED (GLCNAC. . . .) (POTENTIAL).

CC CARBOHYD 542 542 N-LINKED (GLCNAC. . . .) (POTENTIAL).

CC CARBOHYD 558 558 N-LINKED (GLCNAC. . . .) (POTENTIAL).

CC CARBOHYD 578 578 N-LINKED (GLCNAC. . . .) (POTENTIAL).

CC CARBOHYD 627 627 N-LINKED (GLCNAC. . . .) (POTENTIAL).

CC CARBOHYD 649 649 N-LINKED (GLCNAC. . . .) (POTENTIAL).

CC NON_TER 737 737

SQ SEQUENCE 737 AA; 81691 MW; 67DFAE11854122F2 CRC64;

Query Match 77.5%; Score 1158; DB 1; Length 737;

Best Local Similarity 75.6%; Pred. No. 1.6e-92;

Matches 195; Conservative 23; Mismatches 36; Indels 4; Gaps 2;

QY 1 GAKQNVQLINTNGSHLNTALNCNDLSLNTGLAGLFYHHKFNSSGCGPERLASCRPLTDF 60

DB 406 GSRQISLINTNGSHLNTALNCNDLSLNTGLAGLFYHHKFNSSGCGPERLASCRPLTDF 465

QY 61 DOGWGPISYANG--SGPDORPCYCHYPPKPGCVIPAKSVGCVYCFPTSPVVVGTTRSG 118

DB 466 RIGWGTLEYETNVTNEEDMRPYCHYPPKPGCVIPAKSVGCVYCFPTSPVVVGTTRSG 525

QY 119 APTYSWGENDTDFVLNNTTRPPLGNFNGCTWMNSTGFTKVCAGPCVIGGAGNNTLH--C 176

DB 526 VPTYSWGENETDFVLNNTTRPPLGNFNGCTWMNSTGFTKVCAGPCVIGGAGNNTLH--C 585

QY 177 PTDCFRKHPDATYLRKAGGFWLTPCLVDYPIRLMHPCTVNTIFKRMVYGVGVEHRL 236

15-JUL-1999 (Rel. 38, Last annotation update)
GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN); MATRIX
PROTEIN (ENVELOPE PROTEIN M); MAJOR ENVELOPE PROTEIN E; NONSTRUCTURAL
PROTEINS NS1 AND NS2] (FRAGMENT).
OS Hepatitis C virus (isolate HC-J5) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92230232; PubMed=1314459;
RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
RA Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
RT "Full-length sequence of a hepatitis C virus genome having poor
RT homology to reported isolates: comparative study of four distinct
RT genotypes.";
RL Virology 188:331-341(1992).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D10075; BAA00969.1; -
CC INTERPRO: IPR002519; -
CC INTERPRO: IPR002521; -
CC INTERPRO: IPR002522; -
CC INTERPRO: IPR002531; -
CC PFAM: PF01560; HCV_NSI; 1.
CC PFAM: PF01543; HCV_capsid; 1.
CC PFAM: PF01542; HCV_core; 1.
CC PFAM: PF01539; HCV_env; 1.
CC Polyprotein; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural protein.
FT INIT_MET 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 733
FT CHAIN 734 >737
FT TRANSMEM 347 369
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 234 234
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 477 477
FT CARBOHYD 534 534
FT CARBOHYD 542 542
FT CARBOHYD 558 558
FT CARBOHYD 578 578
FT CARBOHYD 627 627
FT CARBOHYD 649 649
FT NON_TER 737 737
SQ SEQUENCE 737 AA; 81207 MW; 3AF699D82AD501B1 CRC64;

Query Match 76.8%; Score 1147; DB 1; Length 737;
Best Local Similarity 75.6%; Pred. No. 1.4e-91;
Matches 195; Conservative 26; Mismatches 33; Indels 4; Gaps 3;

QY 1 GAKONVOLINTNGSWHLNSTALNCNDLNTGLWAGLFYHKKFNSSGCPRLASCRPLTDF 60
DB 406 GSKQNLQINTNGSWHLNSTALNCNDLNTGLWAGLFYHKKFNSSGCPRLASCRSIEAF 465
QY 61 DQGWGPIYA-NGSGP-DORPYCWHYPKPKGIVPAKSYCGVPYCTPSPVVGVTDRSG 118
DB 466 RICWGTQLQYEDNVNTPEDMRPYCWHYPKPKGIVPAKSYCGVPYCTPSPVVGVTDRSG 525
QY 119 APTYSWGENDTDFVNLNTRPPLGNWFGCTWNNSTGCTKVCGAPPCVIGAGN--NTLHC 176
DB 526 VPTYTWGENETDFVNLNTRPPLGNWFGCTWNNSTGCTKVCGAPPCVIGAGN--NTLHC 585
QY 177 PTDGFRKHPDAYSROGSGFWITPRCLVDYPRYLWHPCTINTVIFKIRMYGCVGVEHRL 236
DB 586 PTDGFRKHPDAYSROGSGFWITPRCLVDYPRYLWHPCTINTVIFKIRMYGCVGVEHRL 645
QY 237 AACNWTGRGCDLEDRDR 254
DB 646 AACNWTGRGCDLEDRDR 663

RESULT 11
POLG_HCVH4 STANDARD; PRT; 520 AA.
ID POLG_HCVH4
AC Q01404;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN) (P22);
DE ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35); ENVELOPE GLYCOPROTEIN E2
DE (GP68) (GP70) (NS1)] (FRAGMENT).
DE Hepatitis C virus (isolate HCV-476) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93019030; PubMed=1383400;
RA Abe K., Inchausti G., Fujisawa K.;
RT "Genomic characterization and mutation rate of hepatitis C virus
RT isolated from a patient who contracted hepatitis during an epidemic
RT of non-A, non-B hepatitis in Japan.";
RL J. Gen. Virol. 73:2725-2729(1992).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D10688; BAA01530.1; -
CC INTERPRO: IPR002519; -
CC INTERPRO: IPR002521; -
CC INTERPRO: IPR002522; -
CC INTERPRO: IPR002531; -
CC PFAM: PF01560; HCV_NSI; 1.
CC PFAM: PF01543; HCV_capsid; 1.
CC PFAM: PF01542; HCV_core; 1.
CC PFAM: PF01539; HCV_env; 1.
CC Polyprotein; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural protein.
FT INIT_MET 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 >520
NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
MAJOR ENVELOPE PROTEIN E (POTENTIAL).
MATRIX PROTEIN C (POTENTIAL).
CAPSID PROTEIN C (POTENTIAL).
CELLULAR AMINOPEPTIDASE.

FT TRANSMEM 347 369 BY SIMILARITY.
 FT CARBOHYD 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 418 418 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 424 424 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 431 431 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 449 449 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT NON_TER 520 520 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 520 AA; 56499 MW; AAL35246CF20D525 CRC64;
 Query Match 34.2%; Score 511; DB 1; Length 520;
 Best Local Similarity 73.7%; Pred. No. 6.2e-37;
 Matches 84; Conservative 11; Mismatches 19; Indels 0; Gaps 0;
 QY 1 GAKQNVQLINTNGSWHLNSTALNCNDSLTGWLGLFYHHKFNSSGCPERLASCRPLTDF 60
 Db 407 GPSKIQIOLINTNGSWHLNRTALNCNDSFKTGLAALFYVHKFNASGCPERHMASCRPIDKF 466
 QY 61 DOGWGPISYANGSGDQRPYCHWHPKPCGIVPAKSVCGPVYCFPTSPVVVGT 114
 Db 467 DOGWGPVITYAEPSISEQRPYCHWYAPRCGTIPASEVCGPVYCFPTSPVVVGT 520
 RESULT 12
 ID POLG_HCVHK STANDARD; PRT; 520 AA.
 AC Q01403;
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN) (P22);
 DE ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35); ENVELOPE GLYCOPROTEIN E2
 DE (GP68) (GP70) (NS1)] (FRAGMENT).
 OS Hepatitis C virus (isolate HCV-RF) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 CC Hepacivirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93019030; PubMed=1383400;
 RA Abe K., Inchauspe G., Fujisawa K.;
 RT "Genomic characterization and mutation rate of hepatitis C virus
 RT isolated from a patient who contracted hepatitis during an epidemic
 RT of non-A, non-B hepatitis in Japan."
 RL J. Gen. Virol. 73:2725-2729(1992).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; D10687; BAA01529.1;
 DR PTR; J01925; J01925;
 DR INTERPRO; IPR002519;
 DR INTERPRO; IPR002521;
 DR INTERPRO; IPR002522;
 DR PFAM; PF01560; HCV_NSI.1;
 DR PFAM; PF01543; HCV_capsid.1;
 DR PFAM; PF01542; HCV_core.1;
 DR PFAM; PF01539; HCV_env.1;
 KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;
 KW Transmembrane; Nonstructural protein.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE

FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 384 >520 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT TRANSMEM 347 369 BY SIMILARITY.
 FT CARBOHYD 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 418 418 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 424 424 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 431 431 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 449 449 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT NON_TER 520 520 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 520 AA; 56476 MW; 1D2BD0A6FF27349B CRC64;
 Query Match 34.2%; Score 511; DB 1; Length 520;
 Best Local Similarity 73.7%; Pred. No. 6.2e-37;
 Matches 84; Conservative 11; Mismatches 19; Indels 0; Gaps 0;
 QY 1 GAKQNVQLINTNGSWHLNSTALNCNDSLTGWLGLFYHHKFNSSGCPERLASCRPLTDF 60
 Db 407 GPSKIQIOLINTNGSWHLNRTALNCNDSFKTGLAALFYVHKFNASGCPERHMASCRPIDKF 466
 QY 61 DOGWGPISYANGSGDQRPYCHWHPKPCGIVPAKSVCGPVYCFPTSPVVVGT 114
 Db 467 DOGWGPVITYAEPSISEQRPYCHWYAPRCGTIPASEVCGPVYCFPTSPVVVGT 520
 RESULT 13
 ID POLG_HCVJ2 STANDARD; PRT; 513 AA.
 AC P27959;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN) (P22);
 DE ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35); ENVELOPE GLYCOPROTEIN E2
 DE (GP68) (GP70) (NS1)] (FRAGMENT).
 OS Hepatitis C virus (isolate HCV-J2) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 CC Hepacivirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230232; PubMed=1314459;
 RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Mizuka H.,
 RA Tanaka T., Fukuda S., Tada F., Mishiro S.;
 RT "Full-length sequence of a hepatitis C virus genome having poor
 RT homology to reported isolates: comparative study of four distinct
 RT genotypes."
 RL Virology 188:331-341(1992).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; D10074; BAA00968.1;
 DR INTERPRO; IPR002519;
 DR INTERPRO; IPR002521;

```
DR INTERPRO: IPR002522; -
DR INTERPRO: IPR002531; -
DR PFAM: PF01560; HCV_NSI; 1.
DR PFAM: PF01543; HCV_capsid; 1.
DR PFAM: PF01542; HCV_core; 1.
DR PFAM: PF01539; HCV_env; 1.
DR PFAM: PF01539; HCV_core; 1.
DR Polyprotein; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural protein.
FT INIT_MET 1
FT CHAIN 1
FT CHAIN 115
FT CHAIN 116
FT CHAIN 191
FT CHAIN 192
FT CHAIN 383
FT CHAIN 384
FT CHAIN >513
FT TRANSMEM 347
FT CARBOHYD 369
FT CARBOHYD 196
FT CARBOHYD 209
FT CARBOHYD 233
FT CARBOHYD 234
FT CARBOHYD 234
FT CARBOHYD 250
FT CARBOHYD 305
FT CARBOHYD 417
FT CARBOHYD 423
FT CARBOHYD 430
FT CARBOHYD 448
FT CARBOHYD 513
FT NON_TER 513
SQ SEQUENCE 513 AA; 55704 MW; 943F31E3514CDEF3 CRC64;

Query Match 31.8%; Score 475; DB 1; Length 513;
Best Local Similarity 73.1%; Pred. No. 7.7e-34;
Matches 79; Conservative 9; Mismatches 20; Indels 0; Gaps 0;

QY 1 GAKQNVOLINTNGSWHLNSTALNCNDSLNTGWLGLFYHHKNSGCPERLASCRPLTDF 60
DB 406 GASQKTLINTNGSWHLNSTALNCNDSLNTGFLAALFYTHKFNASGCPERLASCRSDG 465

QY 61 DQGWGPISVANGSGPDRPCYHHPKPGIVPAKSVCGPYVCFTFPSP 108
DB 466 DQGWGPITTEPGDSQDKPCYCHYAPQRCVSVAADVCYVCFTFPSP 513

RESULT 14
POLG_HCVTH STANDARD; PRT; 321 AA.
AC P27957;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: MATRIX PROTEIN (ENVELOPE PROTEIN M);
DE MAJOR ENVELOPE PROTEIN E; NONSTRUCTURAL PROTEIN NS1] (FRAGMENT).
OS Hepatitis C virus (isolate TH) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE-91112009; PubMed=1846505;
RA Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J.,
RA Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M.,
RA Han J.H.;
RT "Variable and hypervariable domains are found in the regions of HCV
RT corresponding to the flavivirus envelope and NS1 proteins and the
RT pestivirus envelope glycoproteins."
RL Virology 180:842-848(1991).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X53136; CAA37296.1; -
```

```
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X53134; CAA37294.1; -
CC INTERPRO: IPR002519; -
CC INTERPRO: IPR002521; -
CC INTERPRO: IPR002531; -
CC PFAM: PF01560; HCV_NSI; 1.
CC PFAM: PF01542; HCV_core; 1.
CC PFAM: PF01539; HCV_env; 1.
CC Polyprotein; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural protein.
FT NON_TER 1
FT CHAIN <1
FT CHAIN 75
FT CHAIN 76
FT CHAIN 267
FT CHAIN >321
FT CARBOHYD 80
FT CARBOHYD 93
FT CARBOHYD 118
FT CARBOHYD 189
FT CARBOHYD 301
FT CARBOHYD 307
FT CARBOHYD 314
FT CARBOHYD 321
FT NON_TER 321
SQ SEQUENCE 321 AA; 34074 MW; B2E883F521C3B520 CRC64;

Query Match 11.6%; Score 173; DB 1; Length 321;
Best Local Similarity 90.6%; Pred. No. 5e-08;
Matches 29; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAKQNVOLINTNGSWHLNSTALNCNDSLNTGWL 32
DB 290 GARQNIQLINTNGSWHLNSTALNCNDSLNTGWL 321

RESULT 15
POLG_HCVEO STANDARD; PRT; 138 AA.
AC P27953;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35);
DE ENVELOPE GLYCOPROTEIN E2 (GP68) (GP70) (NS1)] (FRAGMENT).
OS Hepatitis C virus (isolate EC10) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE-91112009; PubMed=1846505;
RA Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J.,
RA Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M.,
RA Han J.H.;
RT "Variable and hypervariable domains are found in the regions of HCV
RT corresponding to the flavivirus envelope and NS1 proteins and the
RT pestivirus envelope glycoproteins."
RL Virology 180:842-848(1991).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X53136; CAA37296.1; -
```

DR INTERPRO: IPR002519; -
 DR INTERPRO: IPR002531; -
 DR PFAM: PF01560; HCV_NS1; 1.
 DR PFAM: PF01539; HCV_env; 1.
 KW Polypeptide; Glycoprotein; Coat protein; Envelope protein;
 KW Transmembrane; Nonstructural protein.
 FT NON_TER 1 1
 FT CHAIN <1 84 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 85 >138 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CARBOHYD 6 6 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 118 118 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 124 124 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 131 131 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT NON_TER 138 138
 SQ SEQUENCE 138 AA: 14781 MW: 14781 CD3FOA962DEABIAD CRC64;

Query Match 11.4%; Score 171; DB 1; Length 138;
 Best Local Similarity 90.6%; Pred. No. 3e-08;
 Matches 29; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 GAKONVOLINTNGSWHLNSTALNCNDSINTGW 32
 Db 107 GAKONIQLINTNGSWHLNSTALNCNDSINTGW 138

Search completed: March 6, 2001, 11:53:49
 Job time: 186 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 11:55:05 ; Search time 75.33 Seconds
(without alignments)
395.205 Million cell updates/sec

Title: US-09-407-430-3
Perfect score: 1494

Sequence: 1 GAKQNVQLINTNGSWHLNST.....LEAACNWTGRCDELDLDR 254

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL15.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1494	100.0	3011	12	Q91FE5
2	1438	96.3	3011	12	Q03463
3	1426	95.4	778	12	Q04184
4	1418	94.9	3011	12	Q36579
5	1418	94.9	3011	12	Q36608
6	1414	94.6	778	12	Q04185
7	1414	94.6	3011	12	Q36610
8	1393	93.2	3011	12	Q36609
9	1367	91.5	403	12	Q9PX22
10	1304	87.3	3010	12	Q9J3H0
11	1291	86.4	3011	12	Q81754
12	1285	86.0	3010	12	Q81760
13	1284	85.9	3010	12	Q9Q1X6
14	1284	85.9	3010	12	Q9Q1X5
15	1283	85.9	3010	12	Q9J3G4
16	1282	85.8	3010	12	Q9J3G4
17	1282	85.8	3010	12	Q9Q1X6
18	1281	85.7	3010	12	Q9Q1X6
19	1278	85.5	3010	12	Q9J3G1

20	1277.5	85.5	3014	12	O91936	hepatitis c
21	1276	85.4	3010	12	Q02828	h genome po
22	1276	85.4	3010	12	Q9J3G9	hepatitis c
23	1275	85.3	3010	12	Q81989	hepatitis c
24	1275	85.3	3010	12	Q9J3H9	hepatitis c
25	1275	85.3	3010	12	Q9J3H1	hepatitis c
26	1273	85.2	3010	12	Q9Q1X4	hepatitis c
27	1272	85.1	640	12	Q68966	hepatitis c
28	1272	85.1	3010	12	P90193	hepatitis c
29	1272	85.1	3010	12	P90194	hepatitis c
30	1272	85.1	3010	12	Q9Q1Y9	hepatitis c
31	1272	85.1	3010	12	Q9J3H8	hepatitis c
32	1271	85.1	3010	12	Q9Q1X3	hepatitis c
33	1271	85.1	3010	12	Q9J3G0	hepatitis c
34	1269	84.9	3010	12	Q02829	hepatitis c
35	1269	84.9	3010	12	Q02829	h genome po
36	1269	84.9	3010	12	Q9Q1G1	hepatitis c
37	1268	84.9	3010	12	Q9Q1Z0	hepatitis c
38	1268	84.9	3010	12	Q9J3I1	hepatitis c
39	1268	84.9	3010	12	Q9J3G2	hepatitis c
40	1267	84.8	3010	12	Q9J3H7	hepatitis c
41	1266	84.7	3010	12	Q68533	hepatitis c
42	1266	84.7	3010	12	Q9J3H3	hepatitis c
43	1266	84.7	3010	12	Q9J3G3	hepatitis c
44	1265.5	84.7	3015	12	Q9WPH5	hepatitis c
45	1265	84.7	3010	12	O93077	hepatitis c

ALIGNMENTS

RESULT 1

Q91FE5
ID Q91FE5 PRELIMINARY; PRT; 3011 AA.
AC Q91FE5
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE POLYPROTEIN.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Lanford R.E., Lee H., Chavez D., Guerra B., Brasky K.;
RT "An Infectious Clone of the HCV-1 Prototype Sequence."
RL Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL: AF271632; AAF81759.1; -
SQ SEQUENCE 3011 AA; 327126 MW; 2489CE74AC864E58 CRC64;

Query Match 100.0%; Score 1494; DB 12; Length 3011;
Best Local Similarity 100.0%; Pred. No. 3.6e-139;
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gap 0;

Qy	1	GAKQNVQLINTNGSWHLNSTALNCNDSLTGWLACLFYHHKFNSSGCCPERLASCRPLTDF	6
Db	406	GAKQNVQLINTNGSWHLNSTALNCNDSLTGWLACLFYHHKFNSSGCCPERLASCRPLTDF	45
Qy	61	DQGWGPISYANGSGDPORPCYCHWPPKPGCIYPAKSVCGPYCYCTPSPVVGTTDRSGAP	100
Db	466	DQGWGPISYANGSGDPORPCYCHWPPKPGCIYPAKSVCGPYCYCTPSPVVGTTDRSGAP	526
Qy	121	TYSWGENDTDFVLNTRPPLGNFECCTWMNSTGFTKVCAGPCVIGAGNNTLHCPTDC	180
Db	536	TYSWGENDTDFVLNTRPPLGNFECCTWMNSTGFTKVCAGPCVIGAGNNTLHCPTDC	58
Qy	181	FRKHPDATYSRCGGSPWITPRCLVDYPRYLWHYPCTINYTIKIRMYVGVGVEHLEAACN	24
Db	586	FRKHPDATYSRCGGSPWITPRCLVDYPRYLWHYPCTINYTIKIRMYVGVGVEHLEAACN	64
Qy	241	WTRGCRDLEDRDR	254

0.

```

QY 61 DOGWGPISYANGSGDQRPYCHWYPPKPGIIVPAKSVCGPVYCFTPSPVVVGTDRSGAP 120
Db 466 AOGWGPISYANGSGGLDERPYCHWYPPRPGIIVPAKSVCGPVYCFTPSPVVVGTDRSGAP 525
QY 121 TYSWGENDTDFVNLNTRPPLGNWFGCTWMNSTGFTKVCAPPVCGVIGGAGNNTLHCPTDC 180
Db 526 TYSWGANDDTDFVNLNTRPPLGNWFGCTWMNSTGFTKVCAPPVCGVIGGAGNNTLHCPTDC 585
QY 181 FRKHDPATYSRCGSGPWITPRCLVDYPRLWHYPCTINTYIFKIRMYVGGVHEHLEAACN 240
Db 586 FRKHDPATYSRCGSGPWITPRCWDYPRLWHYPCTINTYIFKIRMYVGGVHEHLEAACN 645
QY 241 WTRGERCDLEDRDR 254
Db 646 WTRGERCDLEDRDR 659

RESULT 4
Q36579 ID O36579 PRELIMINARY: PRT: 3011 AA.
AC O36579;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE POLYPROTEIN.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H77;
RX MEDLINE=97373636; PubMed=9228008;
RA Kolykhalov A.A., Agapov E.V., Blight K.J., Mihalik K., Feinstone S.M.,
RA Rice C.M.;
RT "Transmission of hepatitis C by intrahepatic inoculation with
RT transcribed RNA.";
RL Science 277:570-574(1997).
DR EMBL: AF009606; AAB66324.1; -
DR INTERPRO: IPR000745; -
DR INTERPRO: IPR001410; -
DR INTERPRO: IPR001490; -
DR INTERPRO: IPR002166; -
DR INTERPRO: IPR002518; -
DR INTERPRO: IPR002519; -
DR INTERPRO: IPR002521; -
DR INTERPRO: IPR002522; -
DR INTERPRO: IPR002531; -
DR INTERPRO: IPR002868; -
DR PFAM: PF00998; HCV_RdRp; 1.
DR PFAM: PF01001; HCV_NS4b; 1.
DR PFAM: PF01006; HCV_NS4a; 1.
DR PFAM: PF01506; HCV_NS5a; 1.
DR PFAM: PF01538; HCV_NS2; 1.
DR PFAM: PF01539; HCV_env; 1.
DR PFAM: PF01542; HCV_core; 1.
DR PFAM: PF01543; HCV_capsid; 1.
DR PFAM: PF01560; HCV_NSI; 1.
DR PRODOM: PD186062; -; 1.
KW Polyprotein.
SQ SEQUENCE 3011 AA; 327184 MW; E2E0E809C63C1B9 CRC64;

```

```

Query Match 94.9%; Score 1418; DB 12; Length 3011;
Best Local Similarity 94.1%; Pred. No. 1.2e-131;
Matches 239; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 1 GAKQNVQLINTGSHWLNSTALNCNDSLNTGLAGLFYHHKFNSSGCCPERLASCRPLTDF 60
Db 406 GAKQNLINTGSHWLNSTALNCNDSLNTGLAGLFYHHKFNSSGCCPERLASCRPLTDF 465
QY 61 DOGWGPISYANGSGDQRPYCHWYPPKPGIIVPAKSVCGPVYCFTPSPVVVGTDRSGAP 120
Db 466 AOGWGPISYANGSGGLDERPYCHWYPPRPGIIVPAKSVCGPVYCFTPSPVVVGTDRSGAP 525

```

```

Db 466 AOGWGPISYANGSGGLDERPYCHWYPPRPGIIVPAKSVCGPVYCFTPSPVVVGTDRSGAP 525
QY 121 TYSWGENDTDFVNLNTRPPLGNWFGCTWMNSTGFTKVCAPPVCGVIGGAGNNTLHCPTDC 180
Db 526 TYSWGANDDTDFVNLNTRPPLGNWFGCTWMNSTGFTKVCAPPVCGVIGGAGNNTLHCPTDC 585
QY 181 FRKHDPATYSRCGSGPWITPRCLVDYPRLWHYPCTINTYIFKIRMYVGGVHEHLEAACN 240
Db 586 FRKHDPATYSRCGSGPWITPRCWDYPRLWHYPCTINTYIFKIRMYVGGVHEHLEAACN 645
QY 241 WTRGERCDLEDRDR 254
Db 646 WTRGERCDLEDRDR 659

RESULT 5
Q36608 ID O36608 PRELIMINARY: PRT: 3011 AA.
AC O36608;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE POLYPROTEIN.
OS Hepatitis C virus strain H77.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=63746;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H77;
RX MEDLINE=97385173; PubMed=9238047;
RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;
RT "Transcripts from a single full-length cDNA clone of hepatitis C virus
RT are infectious when directly transfected into the liver of a
RT chimpanzee.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:8738-8743(1997).
DR EMBL: AF011751; AAB67036.1; -
DR INTERPRO: IPR000745; -
DR INTERPRO: IPR001410; -
DR INTERPRO: IPR001490; -
DR INTERPRO: IPR002166; -
DR INTERPRO: IPR002518; -
DR INTERPRO: IPR002519; -
DR INTERPRO: IPR002521; -
DR INTERPRO: IPR002522; -
DR INTERPRO: IPR002531; -
DR INTERPRO: IPR002868; -
DR PFAM: PF00998; HCV_RdRp; 1.
DR PFAM: PF01001; HCV_NS4b; 1.
DR PFAM: PF01006; HCV_NS4a; 1.
DR PFAM: PF01506; HCV_NS5a; 1.
DR PFAM: PF01538; HCV_NS2; 1.
DR PFAM: PF01539; HCV_env; 1.
DR PFAM: PF01542; HCV_core; 1.
DR PFAM: PF01543; HCV_capsid; 1.
DR PFAM: PF01560; HCV_NSI; 1.
DR PRODOM: PD186062; -; 1.
KW Polyprotein.
SQ SEQUENCE 3011 AA; 327114 MW; 0B75E6B81CB5C198 CRC64;

```

```

Query Match 94.9%; Score 1418; DB 12; Length 3011;
Best Local Similarity 94.1%; Pred. No. 1.2e-131;
Matches 239; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 1 GAKQNVQLINTGSHWLNSTALNCNDSLNTGLAGLFYHHKFNSSGCCPERLASCRPLTDF 60
Db 406 GAKQNLINTGSHWLNSTALNCNDSLNTGLAGLFYHHKFNSSGCCPERLASCRPLTDF 465
QY 61 DOGWGPISYANGSGDQRPYCHWYPPKPGIIVPAKSVCGPVYCFTPSPVVVGTDRSGAP 120
Db 466 AOGWGPISYANGSGGLDERPYCHWYPPRPGIIVPAKSVCGPVYCFTPSPVVVGTDRSGAP 525

```

Best Local Similarity 92.1%; Pred. No. 7.1e-132; Matches 234; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 121 TYSWGENDTDFVLLNTRPPLGNWFGCTWMNSTGFTKVCAPPVCGAGNNTLHCPTDC 180
DB 526 TYSWGANDDTDFVLLNTRPPLGNWFGCTWMNSTGFTKVCAPPVCGAGNNTLHCPTDC 585

QY 181 FKHPDATYSRCGSGPWITPRCLVDYPRYLWHYPTCTINVTIFKIRMYVGVGVEHRLAAACN 240
DB 586 FRKHPEATYSRCGSGPWITPRCLVDYPRYLWHYPTCTINVTIFKIRMYVGVGVEHRLAAACN 645

QY 241 WTRGERCDLEDRDR 254
DB 646 WTRGERCDLEDRDR 659

RESULT 6
Q04185 PRELIMINARY; PRT; 778 AA.

AC Q04185; Q81811; PRT; 778 AA.

DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C
(CORE PROTEIN); MATRIX PROTEIN
(ENVELOPE PROTEIN M); MAJOR ENVELOPE PROTEIN E; NONSTRUCTURAL PROTEIN
NS1; NONSTRUCTURAL PROTEIN NS2; NONSTRUCTURAL PROTEIN NS4A;
NONSTRUCTURAL PROTEIN NS4B; HELICASE
(NS3); RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (NS5)] (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE=91195357; PubMed=1849654;
RA Ogata N., Alter H.J., Miller R.H., Purcell R.H.;
RT "Nucleotide sequence and mutation rate of the H strain of hepatitis C
virus";
RL Proc. Natl. Acad. Sci. U.S.A. 88:3392-3396(1991).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
HYPOTHETIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION. NS3
AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA.
CC EMBL; M62382; AAB02128.1; -;
DR INTERPRO: IPR002518; -;
DR INTERPRO: IPR002519; -;
DR INTERPRO: IPR002521; -;
DR INTERPRO: IPR002522; -;
DR INTERPRO: IPR002531; -;
DR PFAM; PF01538; HCV_NS2; 1.
DR PFAM; PF01539; HCV_env; 1.
DR PFAM; PF01542; HCV_core; 1.
DR PFAM; PF01543; HCV_capsid; 1.
DR PFAM; PF01560; HCV_NS1; 1.
KW Polyprotein; Glycoprotein; RNA-directed RNA polymerase; Core protein;
Coat protein; Envelope protein; Helicase; ATP-binding; Transmembrane;
Nonstructural protein.
KW INIT_MET 1 REMOVED FROM CAPSID PROTEIN C BY THE
CELLULAR AMINOPEPTIDASE.
FT CHAIN 2 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 729
FT CHAIN 730 >778
FT CHAIN 737 369
FT TRANSMEM 778 778
FT NON_TER 778 778
SQ SEQUENCE 778 AA; 85530 MW; 6E37E50AB97B6C32 CRC64;

QY 121 TYSWGENDTDFVLLNTRPPLGNWFGCTWMNSTGFTKVCAPPVCGAGNNTLHCPTDC 180
DB 526 TYSWGANDDTDFVLLNTRPPLGNWFGCTWMNSTGFTKVCAPPVCGAGNNTLHCPTDC 585

QY 181 FKHPDATYSRCGSGPWITPRCLVDYPRYLWHYPTCTINVTIFKIRMYVGVGVEHRLAAACN 240
DB 586 FRKHPEATYSRCGSGPWITPRCLVDYPRYLWHYPTCTINVTIFKIRMYVGVGVEHRLAAACN 645

QY 241 WTRGERCDLEDRDR 254
DB 646 WTRGERCDLEDRDR 659

RESULT 7
Q036610 PRELIMINARY; PRT; 3011 AA.

AC Q036610;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE POLYPROTEIN.
OS Hepatitis C virus strain H77.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=63746;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=H77;
RX MEDLINE=97385173; PubMed=9238047;
RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;
RT "Transcripts from a single full-length cDNA clone of hepatitis C virus
are infectious when directly transfected into the liver of a
chimpanzee";
RL Proc. Natl. Acad. Sci. U.S.A. 94:8738-8743(1997).
DR EMBL; AF011753; AAB67038.1; -;
DR HSSP; P27958; 1HEI.
DR INTERPRO: IPR000745; -;
DR INTERPRO: IPR001410; -;
DR INTERPRO: IPR001490; -;
DR INTERPRO: IPR002166; -;
DR INTERPRO: IPR002518; -;
DR INTERPRO: IPR002519; -;
DR INTERPRO: IPR002521; -;
DR INTERPRO: IPR002522; -;
DR INTERPRO: IPR002531; -;
DR INTERPRO: IPR002868; -;
DR PFAM; PF00998; HCV_RdRP; 1.
DR PFAM; PF01001; HCV_NS4b; 1.
DR PFAM; PF01006; HCV_NS4a; 1.
DR PFAM; PF01506; HCV_NS5a; 1.
DR PFAM; PF01538; HCV_NS2; 1.
DR PFAM; PF01539; HCV_env; 1.
DR PFAM; PF01542; HCV_core; 1.
DR PFAM; PF01543; HCV_capsid; 1.
DR PFAM; PF01560; HCV_NS1; 1.
DR PRODOM; PD186062; -; 1.
KW Polyprotein.
SQ SEQUENCE 3011 AA; 327222 MW; 293F91430A0D4067 CRC64;

Query Match 94.6%; Score 1414; DB 12; Length 3011;
Best Local Similarity 93.7%; Pred. No. 3e-131;

Matches 238; Conservative 8; Mismatches 8; Indels 0; Gaps 0;

QY 1 GAKONVQLINTNGSWHLNSTALNCNDSLNTGWLAGLFLYHHKFNSSGCPERLASCRPLTDF 60
 Db 406 GAKQIQLINTNGSWHLNSTALNCNDSLNTGWLAGLFLYHHKFNSSGCPERLASCRRLTDF 465

QY 61 DOGWGPISYANGSGDPDRPYCWHYPPKPGCGIVPAKSVCGPVYCFPSPVVVGTTDRSGAP 120
 Db 466 AOGWGPISYANGSGDLDERPYCWHYPPRPGCGIVPAKSVCGPVYCFPSPVVVGTTDRSGAP 525

QY 121 TYSWGENDTDVFLNNTTRPPLGNWFGCTWMNSTGFTKVCAGPPCVIGGAGNNTLHCPTDC 180
 Db 526 TYSWGANDTDVFLNNTTRPPLGNWFGCTWMNSTGFTKVCAGPPCVIGGAGNNTLHCPTDC 585

QY 181 FRKHDPATYSRCGSGPWITPRCLVDYPYRLWHYPCNTINYTIKIRMYVGGVEHRLAECN 240
 Db 586 FRKHPEATYSRCGSGPWITPRCMVDYPYRLWHYPCNTINYTIKIRMYVGGVEHRLAECN 645

QY 241 WTRGERCDLEDRDR 254
 Db 646 WTRGERCDLEDRDR 659

RESULT 8

Q36609 ID Q36609 PRELIMINARY; PRT; 3011 AA.

AC O36609;

DT 01-JAN-1998 (TrEMBLrel. 05, Created)

DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)

DE 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

DE POLYPROTEIN.

OS Hepatitis C virus strain H77.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OX NCBI_TaxID=63746;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN:H77;

RX MEDLINE=97385173; PubMed=9238047;

RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;

RT "Transcripts from a single full-length cDNA clone of hepatitis C virus are infectious when directly transfectected into the liver of a chimpanzee."

RL Proc. Natl. Acad. Sci. U.S.A. 94:8738-8743(1997).

DR EMBL: AF011752; AAB67037.1; -;

DR HSP: P27958; JHEI.

DR INTERPRO: IPR000745; -;

DR INTERPRO: IPR001410; -;

DR INTERPRO: IPR001490; -;

DR INTERPRO: IPR002166; -;

DR INTERPRO: IPR002518; -;

DR INTERPRO: IPR002519; -;

DR INTERPRO: IPR002521; -;

DR INTERPRO: IPR002522; -;

DR INTERPRO: IPR002531; -;

DR INTERPRO: IPR002868; -;

DR PFAM: PF00998; HCV_RdRP; 1.

DR PFAM: PF01001; HCV_NS4B; 1.

DR PFAM: PF01006; HCV_NS4A; 1.

DR PFAM: PF01506; HCV_NS5A; 1.

DR PFAM: PF01538; HCV_NS2; 1.

DR PFAM: PF01539; HCV_env; 1.

DR PFAM: PF01542; HCV_core; 1.

DR PFAM: PF01543; HCV_core; 1.

DR PFAM: PF01560; HCV_NS1; 1.

DR PRODOM: PD186062; -; 1.

KW Polyprotein.

SQ SEQUENCE 3011 AA; 327262 MW; 10DIC9702CA9B5DC CRC64;

Query Match 93.2%; Score 1393; DB 12; Length 3011;
 Best Local Similarity 92.9%; Pred. No. 3.6e-129;
 Matches 236; Conservative 7; Mismatches 11; Indels 0; Gaps 0;

QY 1 GAKONVQLINTNGSWHLNSTALNCNDSLNTGWLAGLFLYHHKFNSSGCPERLASCRPLTDF 60
 Db 406 GAKQIQLINTNGSWHLNSTALNCNDSLNTGWLAGLFLYHHKFNSSGCPERLASCRRLTDF 465

QY 61 DOGWGPISYANGSGDPDRPYCWHYPPKPGCGIVPAKSVCGPVYCFPSPVVVGTTDRSGAP 120
 Db 466 AOGWGPISYANGSGDLDERPYCWHYPPRPGCGIVPAKSVCGPVYCFPSPVVVGTTDRSGAP 525

QY 121 TYSWGENDTDVFLNNTTRPPLGNWFGCTWMNSTGFTKVCAGPPCVIGGAGNNTLHCPTDC 180
 Db 526 TYSWGANDTDVFLNNTTRPPLGNWFGCTWMNSTGFTKVCAGPPCVIGGAGNNTLHCPTDC 585

QY 181 FRKHDPATYSRCGSGPWITPRCLVDYPYRLWHYPCNTINYTIKIRMYVGGVEHRLAECN 240
 Db 586 FRKHPEATYSRCGSGPWITPRCMVDYPYRLWHYPCNTINYTIKIRMYVGGVEHRLAECN 645

QY 241 WTRGERCDLEDRDR 254
 Db 646 WTRGERCDLEDRDR 659

RESULT 9

Q9PX22 ID Q9PX22 PRELIMINARY; PRT; 403 AA.

AC Q9PX22;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DE 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)

DE E2/NS1 REGION.

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OX NCBI_TaxID=11103;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=92348860; PubMed=1668329;

RA Kramsdoerf D., Porchon C., Brechot C.;

RT "Hepatitis C virus (HCV)-RNA in non-A, non-B chronic hepatitis in France. Nucleotide sequence of a French HCV isolate."

RL J. Hepatol. 13:24-24(1991).

DR INTERPRO: IPR002519; -;

DR INTERPRO: IPR002531; -;

DR PFAM: PF01539; HCV_env; 1.

DR PFAM: PF01560; HCV_NS1; 1.

SQ SEQUENCE 403 AA; 44565 MW; 06278192EAC5B3F1 CRC64;

Query Match 91.5%; Score 1367; DB 12; Length 403;
 Best Local Similarity 87.4%; Pred. No. 1.6e-127;
 Matches 222; Conservative 25; Mismatches 7; Indels 0; Gaps 0;

QY 1 GAKONVQLINTNGSWHLNSTALNCNDSLNTGWLAGLFLYHHKFNSSGCPERLASCRPLTDF 60
 Db 73 GAKQIQLINTNGSWHLNSTALNCNDSLNTGWLAGLFLYHHKFNSSGCPERLASCRPLAD 132

QY 61 DOGWGPISYANGSGDPDRPYCWHYPPKPGCGIVPAKSVCGPVYCFPSPVVVGTTDRSGAP 120
 Db 133 DOGWGPISYANGSGDLDERPYCWHYPPRPGCGIVPAKSVCGPVYCFPSPVVVGTTDRSGAP 192

QY 121 TYSWGENDTDVFLNNTTRPPLGNWFGCTWMNSTGFTKVCAGPPCVIGGAGNNTLHCPTDC 180
 Db 193 TYSWGANDTDVFLNNTTRPPLGNWFGCTWMNSTGFTKVCAGPPCVIGGAGNNTLHCPTDC 252

QY 181 FRKHDPATYSRCGSGPWITPRCLVDYPYRLWHYPCNTINYTIKIRMYVGGVEHRLAECN 240
 Db 253 FRKHPEATYSRCGSGPWITPRCLVDYPYRLWHYPCNTINYTIKIRMYVGGVEHRLAECN 312

QY 241 WTRGERCDLEDRDR 254
 Db 313 WTRGERCDLEDRDR 326

RESULT 10
Q9J3H0 PRELIMINARY; PRT; 3010 AA.
AC Q9J3H0;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE POLYPROTEIN.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MD22;
RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT "Characteristics of hepatitis C viral genome associated with disease
RT progression";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF207763; AAF65953.1;
SQ SEQUENCE 3010 AA; 327098 MW; 737EEF31E3C2B28D CRC64;

Query Match 87.3%; Score 1304; DB 12; Length 3010;
Best Local Similarity 83.9%; Pred. No. 2.4e-120; Indels 0; Gaps
Matches 213; Conservative 19; Mismatches 22;

QY 1 GAKQNVQLINTNGSHLNTALNCNDSLNTGLAGLFYHHKFNSSGCPERLASCRPLTDF 60
DB 406 GASONLQINTNGSHWINTALNCNDSLNTGLAGLFYHHKFNSSGCPERLASCRPLTDF 465
QY 61 DQGWGPISYANGSGDPQRYCWHYPPKPCGIVPAKSCVGVCFPTSPVVGTTDRSGAP 120
DB 466 DQGWGPITTEPNSDQRYCWHYAPRPGCIVPASGVGVCFPTSPVVGTTDRSGVP 525
QY 121 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCAPPVCGVAGNNTLHCPTDC 180
DB 526 TYSWGENETDVLNTRPPLGNWFGCTWMNSTGFTKVCAPPVCGVAGNNTLHCPTDC 585
QY 181 FRKHDPATYSCGSGPWTTPCLVDYPRYLWHYPCNTYTFKIRMYGVGVHRELEAACN 240
DB 586 FRKHDPATYSCGSGPWTTPCLVDYPRYLWHYPCNTYTFKIRMYGVGVHRELEAACN 645
QY 241 WTRGERCDLEDRDR 254
DB 646 WTRGERCDLEDRDR 659

RESULT 11
Q81754 PRELIMINARY; PRT; 3011 AA.
AC Q81754;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE POLYPROTEIN.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HC-G9;
RA Okamoto H.;
RL Submitted (APR-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-HC-G9;
RX MEDLINE=94172337; PubMed=8126459;
RA Okamoto H., Kojima M., Sakamoto M., Iizuka H., Hadiwandono S.,
RA Suwignyo S., Miyakawa Y., Mayumi M.;
RT "The entire nucleotide sequence and classification of a hepatitis C
virus isolate of a novel genotype from an Indonesian patient with

chronic liver disease.";
J. Gen. Virol. 75:629-635(1994).
DR EMBL: D14853; BAA03581.1;
DR HSSE; P26663; LJXP.
DR MEROPS; S29.001;
DR INTERPRO: IPR000745;
DR INTERPRO: IPR001410;
DR INTERPRO: IPR001490;
DR INTERPRO: IPR002186;
DR INTERPRO: IPR002518;
DR INTERPRO: IPR002519;
DR INTERPRO: IPR002521;
DR INTERPRO: IPR002522;
DR INTERPRO: IPR002531;
DR INTERPRO: IPR002868;
DR PFAM: PF00998; HCV_RDRP; 1.
DR PFAM: PF01001; HCV_NS4b; 1.
DR PFAM: PF01006; HCV_NS4a; 1.
DR PFAM: PF01506; HCV_NS5a; 1.
DR PFAM: PF01538; HCV_NS2; 1.
DR PFAM: PF01539; HCV_env; 1.
DR PFAM: PF01542; HCV_core; 1.
DR PFAM: PF01543; HCV_capsid; 1.
DR PFAM: PF01560; HCV_NS1; 1.
DR PRODOM: PD186062; -; 1.
KW Polyprotein. 1 191 CORE.
ET CHAIN 192 383 E1.
FT CHAIN 384 809 E2/NS1.
FT CHAIN 810 1006 NS2.
FT CHAIN 1007 1657 NS3.
FT CHAIN 1658 1972 NS4.
FT CHAIN 1973 3011 NS5.
SQ SEQUENCE 3011 AA; 327213 MW; 9C16C120F4E79268 CRC64;

Query Match 86.4%; Score 1291; DB 12; Length 3011;
Best Local Similarity 83.1%; Pred. No. 4.6e-119; Indels 0; Gaps
Matches 211; Conservative 23; Mismatches 20;

QY 1 GAKQNVQLINTNGSHLNTALNCNDSLNTGLAGLFYHHKFNSSGCPERLASCRPLTDF 60
DB 406 GAKQKIQINTNGSHWINTALNCNDSLNTGLAGLFYHHKFNSSGCPERLASCRPLTDF 465
QY 61 DQGWGPISYANGSGDPQRYCWHYPPKPCGIVPAKSCVGVCFPTSPVVGTTDRSGAP 120
DB 466 DQGWGPITTEPNSDQRYCWHYAPRPGCIVPASGVGVCFPTSPVVGTTDRSGVP 525
QY 121 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCAPPVCGVAGNNTLHCPTDC 180
DB 526 TYRWGANETDVLNTRPPLGNWFGCTWMNSTGFTKVCAPPVCGVAGNNTLHCPTDC 585
QY 181 FRKHDPATYSCGSGPWTTPCLVDYPRYLWHYPCNTYTFKIRMYGVGVHRELEAACN 240
DB 586 FRKHDPATYSCGSGPWTTPCLVDYPRYLWHYPCNTYTFKIRMYGVGVHRELEAACN 645
QY 241 WTRGERCDLEDRDR 254
DB 646 WTRGERCDLEDRDR 659

RESULT 12
Q81760 PRELIMINARY; PRT; 3010 AA.
ID Q81760;
AC Q81760;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE POLYPROTEIN PRECURSOR.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.

OX NCBI_TaxID=11103;
[1]

RP SEQUENCE FROM N.A.

RC STRAIN=HC-C2;

RX MEDLINE=93359897; PubMed=8394876;

RA Wang Y., Okamoto H., Tsuda F., Nagayama R., Tao Q.M., Mishiro S.;
RT "Prevalence, genotypes, and an isolate (HC-C2) of hepatitis C virus in
Chinese patients with liver disease.";

RL J. Med. Virol. 40:254-260(1993).

DR EMBL: D10934; BAA01728.1; -

DR HSSP: P26663; INS3

DR INTERPRO: IPR00745; -

DR INTERPRO: IPR001410; -

DR INTERPRO: IPR001490; -

DR INTERPRO: IPR002166; -

DR INTERPRO: IPR002518; -

DR INTERPRO: IPR002519; -

DR INTERPRO: IPR002521; -

DR INTERPRO: IPR002522; -

DR INTERPRO: IPR002868; -

DR PFAM: PF00998; HCV_RdRp; 1.

DR PFAM: PF01001; HCV_NS4b; 1.

DR PFAM: PF01006; HCV_NS4a; 1.

DR PFAM: PF01506; HCV_NS3a; 1.

DR PFAM: PF01538; HCV_NS2; 1.

DR PFAM: PF01539; HCV_env; 1.

DR PFAM: PF01542; HCV_core; 1.

DR PFAM: PF01543; HCV_capsid; 1.

DR PFAM: PF01560; HCV_NS1; 1.

DR PRODOM: PD186062; -; 1.

KW Polyprotein.

SQ SEQUENCE 3010 AA; 326857 MW; EA7D306A4BA2E224 CRC64;

Query Match

Best Local Similarity 86.0%; Score 1285; DB 12; Length 3010;

Matches 208; Conservative 21; Mismatches 25; Indels 0; Gaps 0;

QY 1 GAKQNVQLINTNGSWHLNSTALNCNDSLNTGWLGLFYLHKKFNSSGCPERLASCRPLTDF 60

DB 406 GPSQKIQLNTNGSWHLNSTALNCNDSLNTGWLGLFYLHKKFNSSGCPERLASCRPLTDF 60

QY 61 DQGWGPISTANGSGDQRPYCHYHPKPCGIVPAKSVCGPVYCFPTSPVVGTTDRSGAP 120

DB 466 DQGWGPISTANGSGDQRPYCHYHPKPCGIVPAKSVCGPVYCFPTSPVVGTTDRSGAP 120

QY 121 TYSWGENDTDFVLLNTRPPLGNWFGCTWNTGTFTKIRMYVGGVHRLEAACN 180

DB 526 TYNWGENETDVLNTRPPLGNWFGCTWNTGTFTKIRMYVGGVHRLEAACN 180

QY 181 FRKHDPATYRCGSGPWITPRCLVDYPYRLWHYPCNTINTYIFKIRMYVGGVHRLEAACN 240

DB 586 FRKHDPATYRCGSGPWITPRCLVDYPYRLWHYPCNTINTYIFKIRMYVGGVHRLEAACN 240

QY 241 WTRGERCDLEDRDR 254

DB 646 WTRGERCDLEDRDR 659

RESULT 13

QYQIY6

ID QYQIY6 PRELIMINARY; PRT; 3010 AA.

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

DE POLYPROTEIN.

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OX NCBI_TaxID=11103;

RP SEQUENCE FROM N.A.

RC STRAIN=MD3-2;

RX MEDLINE=20013325; PubMed=10544098;

RA Nagayama K., Kurosaki M., Enomoto N., Maekawa S.Y., Miyasaka Y.,

RP SEQUENCE FROM N.A.

RC STRAIN=MD3-1;

RX MEDLINE=20013325; PubMed=10544098;

RA Nagayama K., Kurosaki M., Enomoto N., Maekawa S.Y., Miyasaka Y.,

RA Tazawa J.I., Izumi N., Marumo F., Sato C.;

RT "Time-related changes in full-length hepatitis C virus and hepatitis

activity.";

RL Virology 263:244-253(1999).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=MD3-1;

RA Nagayama K., Kurosaki M., Enomoto N., Maekawa S., Miyasaka Y.,

RA Sakamoto N., Fukuma T., Tazawa J., Izumi N., Marumo F., Sato C.;

RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF165049; AAD56184.1; -

DR HSSP: P26663; LJXP

DR INTERPRO: IPR00745; -

DR INTERPRO: IPR001410; -

DR INTERPRO: IPR002166; -

DR INTERPRO: IPR002519; -

DR INTERPRO: IPR002531; -

DR PFAM: PF00998; HCV_RdRp; 1.

DR PFAM: PF01006; HCV_NS4a; 1.

DR PFAM: PF01538; HCV_NS2; 1.

DR PFAM: PF01539; HCV_env; 1.

DR PFAM: PF01560; HCV_NS1; 1.

KW Polyprotein.

SQ SEQUENCE 3010 AA; 327369 MW; 998C7F293EAAEC8D CRC64;

Query Match

Best Local Similarity 85.9%; Score 1284; DB 12; Length 3010;

Matches 210; Conservative 17; Mismatches 27; Indels 0; Gaps 0;

QY 1 GAKQNVQLINTNGSWHLNSTALNCNDSLNTGWLGLFYLHKKFNSSGCPERLASCRPLTDF 60

DB 406 GASQKIQLNTNGSWHLNSTALNCNDSLNTGWLGLFYLHKKFNSSGCPERLASCRPLTDF 60

QY 61 DQGWGPISTANGSGDQRPYCHYHPKPCGIVPAKSVCGPVYCFPTSPVVGTTDRSGAP 120

DB 466 DQGWGPISTANGSGDQRPYCHYHPKPCGIVPAKSVCGPVYCFPTSPVVGTTDRSGAP 120

QY 121 TYSWGENDTDFVLLNTRPPLGNWFGCTWNTGTFTKIRMYVGGVHRLEAACN 180

DB 526 TYSWGENETDVLNTRPPLGNWFGCTWNTGTFTKIRMYVGGVHRLEAACN 180

QY 181 FRKHDPATYRCGSGPWITPRCLVDYPYRLWHYPCNTINTYIFKIRMYVGGVHRLEAACN 240

DB 586 FRKHDPATYRCGSGPWITPRCLVDYPYRLWHYPCNTINTYIFKIRMYVGGVHRLEAACN 240

QY 241 WTRGERCDLEDRDR 254

DB 646 WTRGERCDLEDRDR 659

RESULT 14

QYQIY5

ID QYQIY5 PRELIMINARY; PRT; 3010 AA.

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

DE POLYPROTEIN.

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OX NCBI_TaxID=11103;

RP SEQUENCE FROM N.A.

RC STRAIN=MD3-2;

RX MEDLINE=20013325; PubMed=10544098;

RA Nagayama K., Kurosaki M., Enomoto N., Maekawa S.Y., Miyasaka Y.,

Tue Mar 6 12:02:35 2001

RA Tazawa J.i., Izumi N., Marumo F., Sato C.;
RT "Time-related changes in full-length hepatitis C virus and hepatitis
RL activity";
VL Virology 263:244-253(1999).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-MD3-2;

RA Nagayama K., Kurosaki M., Enomoto N., Maekawa S., Miyasaka Y.,

RA Sakamoto N., Fukuma T., Tazawa J., Izumi N., Marumo F., Sato C.;

PL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF165050; AAD56185.1;

DR HSP; P28663; IJXP.

DR INTERPRO: IPR000745;

DR INTERPRO: IPR001410;

DR INTERPRO: IPR002166;

DR INTERPRO: IPR002518;

DR INTERPRO: IPR002519;

DR INTERPRO: IPR002531;

DR PFAM: PF00998; HCV_RdRP; 1.

DR PFAM: PF01006; HCV_NS4a; 1.

DR PFAM: PF01538; HCV_NS2; 1.

DR PFAM: PF01539; HCV_env; 1.

DR PFAM: PF01560; HCV_NS1; 1.

KW Polyprotein

SQ SEQUENCE 3010 AA; 327431 MW; 15190E3463DEB35 CRC64;

Query Match 85.9%; Score 1284; DB 12; Length 3010;
Best Local Similarity 82.7%; Pred. No. 2.3e-118;
Matches 210; Conservative 17; Mismatches 27; Indels 0; Gaps 0;

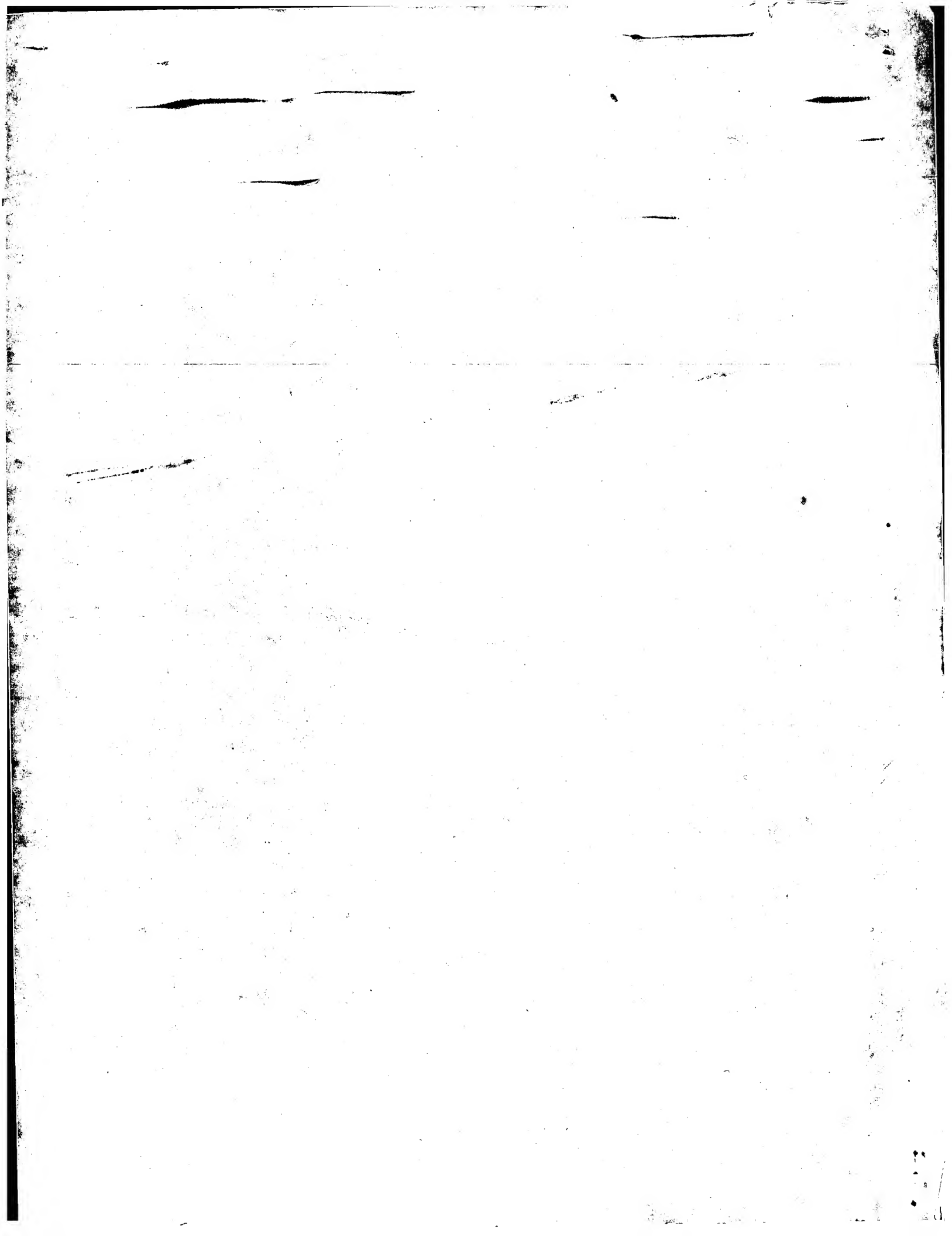
QY 1 GAKQNVQLINTGSHLNTALNCNDLSLNTGLAGLFYHHKFNSSGCPERLASCRPLTDF 60
DB 406 GASQRIQLINTGSHLNTALNCNDLSLNTGLAALFYTHRFNSSGCPERLASCRSIDKF 465
QY 61 DOGWGPIYANGSGDPQRPYCHWHYPPKPGIVPAKSVGPGVYCFPTSPVVVGTTRDSGAP 120
DB 466 DOGWGPIYASPGQDQRPYCHWHYAPKOGIVPASQVCGPVYCFPTSPVVVGTTRDFGVP 525
QY 121 TVSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCAPPVIGGAGNNTLHCPTDC 180
DB 526 TVSWGNETDVLNTRPPOGQNWFGCTWMNGTGFTKTCGGPPCNIGGAGNNTLTCPTDC 585
QY 181 FRKHDPATYSRCGSPWITPRCLVDYPYRLWHYPCTVNTYTFKIRMYVGVGVEHRLAAACN 240
DB 586 FRKHPEATYTKCGSPWLTPRCLVDYPYRLWHYPCTVNTYTFKIRMYVGVGVEHRLAAACN 645
QY 241 WTRGERCDLEDRDR 254
DB 646 WTRGERCDLEDRDR 659

RESULT 15
Q9J3G4 PRELIMINARY; PRT: 3010 AA.
AC Q9J3G4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE POLYPROTEIN.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MD28;
RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT "Characteristics of hepatitis C viral genome associated with disease
RT progression";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
EMBL: AF207769; AAF65959.1;
SQ SEQUENCE 3010 AA; 327318 MW; 0595728AC62464F3 CRC64;

Query Match 85.9%; Score 1283; DB 12; Length 3010;
Best Local Similarity 82.3%; Pred. No. 2.8e-118;
Matches 209; Conservative 20; Mismatches 25; Indels 0; Gaps 0;

QY 1 GAKQNVQLINTGSHLNTALNCNDLSLNTGLAGLFYHHKFNSSGCPERLASCRPLTDF 60
DB 406 GASQRIQLINTGSHLNTALNCNDLSLNTGLAALFYTHRFNSSGCPERLASCRSIDTF 465
QY 61 DOGWGPIYANGSGDPQRPYCHWHYPPKPGIVPAKSVGPGVYCFPTSPVVVGTTRDSGAP 120
DB 466 DOGWGPIYAKSGSPQRPYCHWHYAPRPGVVPASEVCGPVYCFPTSPVVVGTTRDFGAP 525
QY 121 TVSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCAPPVIGGAGNNTLHCPTDC 180
DB 526 TVSWGNETDVLNTRPPOGQNWFGCTWMNGTGFTKTCGGPPCNIGGAGNNTLTCPTDC 585
QY 181 FRKHDPATYSRCGSPWITPRCLVDYPYRLWHYPCTVNTYTFKIRMYVGVGVEHRLAAACN 240
DB 586 FRKHPEATYTKCGSPWLTPRCLVDYPYRLWHYPCTVNTYTFKIRMYVGVGVEHRLAAACN 645
QY 241 WTRGERCDLEDRDR 254
DB 646 WTRGERCDLEDRDR 659

Search completed: March 6, 2001, 11:55:12
Job time: 269 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 11:50:43 ; Search time 75.33 Seconds
(without alignments)
261.396 Million cell updates/sec

Title: US-09-407-430-1
Perfect score: 921
Sequence: 1 MNSKGOYPTQPTYPVQPGN.....VTQRKGNFFMGSDGGYTIW 168

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues 374700

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_15.*

- 1: sp-archaea.*
- 2: sp-bacteria.*
- 3: sp-fungi.*
- 4: sp-human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp-organelle.*
- 9: sp-phage.*
- 10: sp-plant.*
- 11: sp-rodent.*
- 12: sp-virus.*
- 13: sp-vertebrate.*
- 14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	921	100.0	168	4	Q15038	Q15038 homo sapien
2	911	98.9	168	11	O88675	O88675 mus musculus
3	129	14.0	502	4	Q9UG75	Q9UG75 homo sapien
4	126	13.7	856	4	O76022	O76022 homo sapien
5	124	13.5	262	3	O94274	O94274 schizosacch
6	124	13.5	291	10	Q9SKI0	Q9SKI0 arabidopsis
7	124	13.5	494	10	Q93367	Q93367 glycine max
8	124	13.5	1840	10	Q9S258	Q9S258 arabidopsis
9	123.5	13.4	2703	5	Q9VEG7	Q9VEG7 drosophila
10	123.5	13.4	2715	5	O61603	O61603 drosophila
11	123	13.4	625	10	Q99366	Q99366 glycine max
12	121.5	13.2	977	10	Q99368	Q99368 glycine max
13	120.5	13.1	468	5	Q9NF32	Q9NF32 drosophila
14	120.5	13.1	469	5	Q9WSD6	Q9WSD6 drosophila
15	120	13.0	1126	5	Q19371	Q19371 caenorhabdi
16	119	12.9	329	4	Q9NRQ2	Q9NRQ2 homo sapien
17	119	12.9	1541	5	O15837	O15837 leishmania
18	118.5	12.9	239	5	Q9W2R5	Q9W2R5 drosophila
19	118	12.8	1690	5	O77165	O77165 mastigamoeb

20 117.5 12.8 530 10 Q9LYK5
21 117.5 12.8 926 5 Q9W3G1
22 117.5 12.8 1811 5 Q9XZU8
23 117.5 12.8 1887 5 Q9VYX6
24 117.5 12.8 1889 5 Q9XZS2
25 117.5 12.8 1889 5 Q9XZS1
26 117.5 12.8 1889 5 Q9XZU9
27 117 12.7 1862 5 Q20090
28 116.5 12.6 990 4 Q9UG03
29 116.5 12.6 1345 4 Q9P257
30 116.5 12.6 1553 5 Q96452
31 116 12.6 261 4 Q9P112
32 115 12.5 450 5 Q9VYF9
33 114.5 12.4 440 5 Q9VR13
34 114.5 12.4 552 5 Q76861
35 114.5 12.4 574 4 Q08397
36 114 12.4 472 3 Q59907
37 113.5 12.3 414 3 Q94231
38 113.5 12.3 857 3 Q9P7E8
39 113.5 12.3 1039 5 Q9VKG4
40 113 12.3 261 11 P97765
41 113 12.3 306 11 Q35449
42 113 12.3 1605 5 Q96446
43 112 12.2 1668 4 Q15026
44 112 12.2 2971 4 Q9Y5L9
45 111.5 12.1 237 5 Q17242

ALIGNMENTS

RESULT 1

Q15038
ID Q15038; PRELIMINARY; PRT; 168 AA.
AC Q15038;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE KIAA0058 PROTEIN.
GN KIAA0058.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96051398; PubMed=7584044;
RA Nomura N., Nagase T., Miyajima N., Sazuka T., Tanaka A., Sato S.,
RA Seki N., Kavarabayashi Y., Ishikawa K., Tabata S.;
RT "Prediction of the coding sequences of unidentified human genes. II.
RT The coding sequences of 40 new genes (KIAA0041-KIAA0080) deduced by
RT analysis of cDNA clones from human cell line KG-1.";
RL DNA Res. 1:223-229(1994).
DR ENGL; D31767; BAA06545.1;
DR INTERSEQ; IPR000515;
DR PROSITE; PS00402; BPD-TRANSP_INN_MEMBR; UNKNOWN_1.
SQ SEQUENCE 168 AA; 17319 MW; 49F1B6D281E24AAC CRC64;

Query Match 100.0%; Score 921; DB 4; Length 168;
Best Local Similarity 100.0%; Pred No. 6.2e-68;
Matches 168; Conservative 0; Mismatches 0; Indels 0; Saps 0;

QY 1 MNSKGOYPTQPTYPVQPGNPVQPTLHLPAQPPYTDAPPAYSELYRPSFVHPGAATV 60
Db 1 MNSKGOYPTQPTYPVQPGNPVQPTLHLPAQPPYTDAPPAYSELYRPSFVHPGAATV 60
QY 61 MSAAFPGASLYLPMASVAVGLGSTITPMAYYPVGIPYPPGSTVLVEGGYDAGARFGAGA 120
Db 61 MSAAFPGASLYLPMASVAVGLGSTITPMAYYPVGIPYPPGSTVLVEGGYDAGARFGAGA 120
QY 121 TAGNIPPPPGCPNNAQAQLAVMOGANVLVTQRKGNFFMGSDGGYTIW 168
|||||

DT	01-MAY-2000 (TReMBLrel. 13, Last sequence update)	OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
DT	01-OCT-2000 (TReMBLrel. 15, Last annotation update)	CC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
DE	DNA-DIRECTED RNA POLYMERASE (EC 2.7.7.6) II LARGEST CHAIN	CC	Ephydroidea; Drosophilidae; Drosophila.
DE	(EC 2.7.7.6).	OX	NCBI_TaxID=7227;
GN	FAB14.70 OR ATAG35800.	RN	[1]
OS	Arabidopsis thaliana (Mouse-ear cress).	RP	SEQUENCE FROM N.A.
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;	RC	STRAIN=BERKELEY;
OC	Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;	RX	MEDLINE=20196006; PubMed=10731132;
OC	Brassicaceae; Brassicaceae; Arabidopsis.	RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
OX	NCBI_TaxID=3702;	RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA	SEQUENCE FROM N.A.	RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA	Bevan M., Rose M., Hempel S., Entian K.-D., Hoheisel J., Jesse T.,	RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA	Heijnen L., Vos P., Mewes H.W., Mayer K.F.X., Schueller C.;	RA	Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA	Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.	RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RL	SEQUENCE FROM N.A.	RA	Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.	RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.	RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.	RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.	RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.	RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.	RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.	RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.	RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.	RA	Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.	RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.	RA	Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.	RA	Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.	RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.	RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.	RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.	RA	Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.	RA	Merkulov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.	RA	Mout S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.	RA	Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.	RA	Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.	RA	Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.	RA	Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.	RA	Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.	RA	Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.	RA	Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.	RA	Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.	RA	Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.	RA	Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.	RA	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.	RA	"The genome sequence of Drosophila melanogaster."
RT	Science 287:2185-2195(2000).	RT	Science 287:2185-2195(2000).
DR	EMBL; AF003718; AAF55457.1;	DR	EMBL; AF003718; AAF55457.1;
DR	FLYBASE; FBgn0003013; osa.	DR	FLYBASE; FBgn0003013; osa.
DR	INTERPRO; IPR000104;	DR	INTERPRO; IPR000104;
DR	INTERPRO; IPR001606;	DR	INTERPRO; IPR001606;
DR	INTERPRO; IPR002965;	DR	INTERPRO; IPR002965;
DR	PFAM; PF01388; ARID; 1.	DR	PFAM; PF01388; ARID; 1.
DR	PRINTS; PR00308; ANTIFREEZE1.	DR	PRINTS; PR00308; ANTIFREEZE1.
DR	PRINTS; PR01217; PRICHTEXTENS.	DR	PRINTS; PR01217; PRICHTEXTENS.
DR	SEQUENCE 2703 AA; 282894 MW; 27ED5BA0277A6369 CRC64;	DR	SEQUENCE 2703 AA; 282894 MW; 27ED5BA0277A6369 CRC64;
QY	8 PTPQTY-PVOPGPNVYPTLHPQAPYTDAPPAYSELYRPSFVHFGAATVPTMSAAPP 66	QY	8 PTPQTY-PVOPGPNVYPTLHPQAPYTDAPPAYSELYRPSFVHFGAATVPTMSAAPP 66
DB	1673 PTPSISPTSPSYSTPS--YSPTSPAYSTSPGYSPT-SFYSPTSPSYSTSPSYNP 1729	DB	1673 PTPSISPTSPSYSTPS--YSPTSPAYSTSPGYSPT-SFYSPTSPSYSTSPSYNP 1729
QY	67 GASLYLPMQAQSVAVGFLGSTIPMA-----YYPVGPITPPCS-TVLVEGG 109	QY	67 GASLYLPMQAQSVAVGFLGSTIPMA-----YYPVGPITPPCS-TVLVEGG 109
DB	1730 QSAKYSPTSPSYSTSPSYSTSPSYSTSPSYSTSPSYSTSPSYSTSPSYSTSP 1786	DB	1730 QSAKYSPTSPSYSTSPSYSTSPSYSTSPSYSTSPSYSTSPSYSTSPSYSTSP 1786
QY	110 YDAGARFGAGATAGNIPPPPPCCPNAAQAVMOG 144	QY	110 YDAGARFGAGATAGNIPPPPPCCPNAAQAVMOG 144
DB	1787 YSSGASPDYSPSAGYSTPLFGYSPSTGQYTPHEG 1821	DB	1787 YSSGASPDYSPSAGYSTPLFGYSPSTGQYTPHEG 1821
RESULT	9	RESULT	9
Q9VEG7	Q9VEG7 PRELIMINARY; PRT; 2703 AA.	Q9VEG7	Q9VEG7 PRELIMINARY; PRT; 2703 AA.
AC	Q9VEG7; (TReMBLrel. 13, Created)	AC	Q9VEG7; (TReMBLrel. 13, Created)
DT	01-MAY-2000 (TReMBLrel. 13, Last sequence update)	DT	01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT	01-JUN-2000 (TReMBLrel. 14, Last annotation update)	DT	01-JUN-2000 (TReMBLrel. 14, Last annotation update)
DE	OSA PROTEIN.	DE	OSA PROTEIN.
GN	Drosophila melanogaster (Fruit fly).	GN	Drosophila melanogaster (Fruit fly).

```

RESULT 10
ID O61603 PRELIMINARY; PRT: 2715 AA.
AC O61603;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE EYELID.
GN ELD.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97415319; PubMed=9271118;
RA Treisman J.E., Luk A., Rubin G.M., Heberlein U.;
RT "eyelid antagonizes wingless signaling during Drosophila development
RT and has homology to the Bright family of DNA-binding proteins.";
RL Genes Dev. 11:1949-1962(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Treisman J.E., Luk A., Rubin G.M., Heberlein U.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF053091; AAC06254.1; -
DR FLYBASE; FBgn015384; eld.
DR INTERPRO; IPR001606; -.
DR PFAM; PF01388; ARID; 1.
SQ SEQUENCE 2715 AA; 283935 MW; 6424C5FF36E93702 CRC64;

Query Match 13.4%; Score 123.5; DB 5; Length 2715;
Best Local Similarity 31.0%; Pred. NO. 0.078;
Matches 53; Conservative 4; Mismatches 81; Indels 33; Gaps

Qy 1 MNSKGQVPTQPTYPVQ-----PGNPVYPQTLHPQAPP--YTDAPPAYSELYRPSFVHP 53
Db 636 MGGFGQSGAGGYPPQPOQYPPGN---YPPR---PQYPPGAYATGPP----- 677

Qy 54 GAATVPTMSAFFGASLYLPMQAQSVAVPLGSLTPMAYYPVGPIYPGSGTVLVEGGYDAG 113
Db 678 ---PPPTSQAGAGGASMPGSAAGGYPGRGMPNHTGQYPPYQWVPSPQQTVPGAGPG 734

Qy 114 ARFCAGATAGNIPPPP--PGCPPNNAQLAVMQGANLV--TQRKGNFFMGGS 161
Db 735 AMVGNHVQGRKTPPPVVGPPPGQSGSPRPLNLYKQLHQHKGV--GGS 783

RESULT 11
ID Q99366 PRELIMINARY; PRT: 625 AA.
AC Q99366;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE DNA-DIRECTED RNA POLYMERASE (FRAGMENT).
GN RPL1-B1 OR GENE B1.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Fabales; Fabaceae; Papilionoideae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CV. WAYNE;
RA Dietrich M.A.;
RL Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 364-625 FROM N.A.
RX STRAIN=CV. WAYNE;
RP MEDLINE=91355869; PubMed=2103447;

```

Query Match 13.2%; Score 121.5; DB 10; Length 977;
Best Local Similarity 26.9%; Pred. No. 0.039;
Matches 39; Conservative 16; Mismatches 65; Indels 25; Gaps 5;

QY 3 SKGQYPTQPTYPVOPPG-NPVYPTQTLHLPOAPYTDAPPAYSELYRPSFVHFGAAVPTM 61
DB 756 SPQYSPTSPTVSPSSPGYSTSP--AYSPTSPSYSTSPAYSPT-SPAYSSTSPAYSPTS 812
QY 62 SAAPFGASLYLPMQAQSVAV-----GPLGSTIPMAYYVGVGDIYPPGSGVNLVEGGYDAG 113
DB 813 PSYSPTSPAYSPSYSPSYLSPSYSPSYSTSPSYSTSPSYSTSPSYSTSPSYSTSPSYSTSP 862
QY 114 ARFGAGATAGNIPPPPPGCPNAAQ 138
DB 863 --SYSPSTSPSYSTSPSYSPSYNPOSAS 884

RESULT 13
Q9NF32 PRELIMINARY; PRT; 468 AA.
AC Q9NF32;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE EG:BACR19J1.3 PROTEIN.
GN EG:BACR19J1.3
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Borkova D., Minana B., Kafatos F.C.;
RT "Sequencing the distal x chromosome of Drosophila melanogaster.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Benos P.;
RA Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RL EMBL; AL132792; CAB60032.1;
DR EMBL; AL132792; CAB60032.1;
SQ SEQUENCE 468 AA; 46357 MW; CCB0B4FED5FBC5F4 CRC64;

Query Match 13.1%; Score 120.5; DB 5; Length 468;
Best Local Similarity 28.4%; Pred. No. 0.022;
Matches 42; Conservative 11; Mismatches 52; Indels 43; Gaps 8;

QY 11 PTYPVQPPGNPVYPTQTLHLP-----QAPPYTDAPPAY-----SELYRPSF-----VHP 53
DB 114 PAYSASNPAGGARP-NMHEPPPAYHAPNYGAAPPNYGAATGSNVHQPYSGVPGATYYP 172
QY 54 GAATVPTMSAAPPAGSALYLPMAQSVAVG-----PLGSTIPMAYYVGVGDIYPPGSTV 104
DB 173 AAGPGSGYSPNIPAGATYYPFAGVPMGGYHPAAAPPAGAT-----YYQAGSALPPGATY 228
QY 105 LVEGGYDA-----GARFGAGATAGNI 125
DB 229 -----YSAPPOQSSSGLGFGTLLAGGL 251

RESULT 14
Q9W5D6 PRELIMINARY; PRT; 469 AA.
AC Q9W5D6;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
DE CG5273 PROTEIN.
GN CG5273.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;

pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
[1]
SEQUENCE FROM N.A.
STRAIN=BERKELEY;
MEDLINE=20196006; PubMed=10731132;
ADAMS M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
Sutton G.G., Wortman J.E., Richards S., Ashburner M., Henderson S.N.,
Sutton G.G., Rogers Y.-H.C., Blazek R.G., Champe C.R., Miklos G.L.G.,
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Baldwin D.,
Abril J.F., Agbavani A., An H.-J., Andrews-Pfannkuch C., Beasley E.M.,
Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
Cherry J.M., Cawley S., Dahle C., Daves A.D., Dew I., Dietz S.M.,
de Pablos B., Delcher A., Deng Z., Dugan-Rocha S., Dunkov B.C., Dunn P.,
Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Fleischmann W.,
Dorbin K.J., Evangelista C.C., Ferraz C., Ferrer J., Glasser K.,
Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Harris M.,
Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
Hostin D., Houston K.A., Howland T.J., Ke Z., Kennison J.A., Ketchum K.A.,
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
Kimmel B.E., Kodira C.D., Kraft C., Krawitz S., Kulp D., Lai Z.,
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
Mout S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
Reibert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J., Q.A.,
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhou X., Zhu S.,
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AF003418; AAF45545.1;
DR FLYBASE; FBgn0029534; CG5273.
SQ SEQUENCE 469 AA; 46488 MW; 693F759B4ED964A0 CRC64;

Query Match 13.1%; Score 120.5; DB 5; Length 469;
Best Local Similarity 28.4%; Pred. No. 0.022;
Matches 42; Conservative 11; Mismatches 52; Indels 43; Gaps 8;

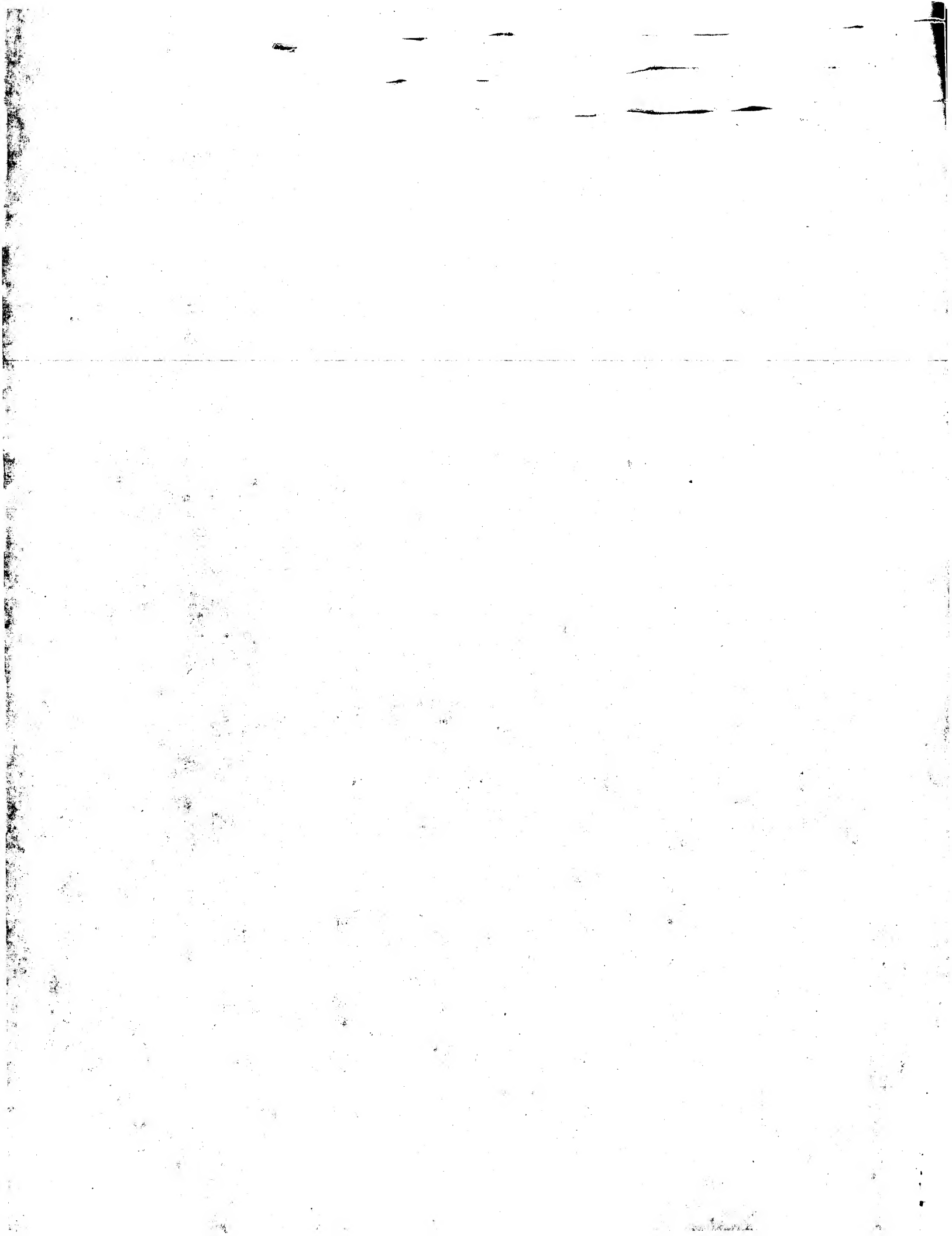
QY 11 PTYPVQPPGNPVYPTQTLHLP-----QAPPYTDAPPAY-----SELYRPSF-----VHP 53
DB 115 PAYSASNPAGGARP-NMHEPPPAYHAPNYGAAPPNYGAATGSNVHQPYSGVPGATYYP 173
QY 54 GAATVPTMSAAPPAGSALYLPMAQSVAVG-----PLGSTIPMAYYVGVGDIYPPGSTV 104
DB 174 AAGPGSGYSPNIPAGATYYPFAGVPMGGYHPAAAPPAGAT-----YYQAGSALPPGATY 229
QY 105 LVEGGYDA-----GARFGAGATAGNI 125
DB 230 -----YSAPPOQSSSGLGFGTLLAGGL 252

RESULT 15
Q19371 PRELIMINARY; PRT; 1126 AA.
AC Q19371;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)

DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE F12F6.6 PROTEIN.
GN F12F6.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Kershaw J.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA Smaldon N., Smith A., Sonnenhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.*;
RL Nature 368:32-38(1994).
DR EMBL: Z73425; CAA97788.1; -;
DR INTERPRO: IPR000255; -;
DR PROSITE: PS00012; PHOSPHOPANTHETHEINE; UNKNOWN_1.
SQ SEQUENCE 1126 AA; 121122 MW; 67C1FADE3D3F3EB3 CRC64;

Query Match 13.0%; Score 120; DB 5; Length 1126;
Best Local Similarity 26.4%; Pred. No. 0.061; Mismatches 67; Indels 74; Gaps 11;
Matches 55; Conservative 12; Mismatches 67; Indels 74; Gaps 11;
QY 1 MNSKGQYPTQTYVPQGNVVPQTLHLPQA-----PPYTD 37
DB 151 MNOGAHLPPQPHQHQP--TLRPQIPGPQAGGSSFSQAGAPTSQNVQYVGGP--SS 206
QY 38 APPAYSELYRPSEVHPGAATVPTMSAAF--PGASLYLPMAQSVAVGVLGSTIPMAY---- 91
DB 207 APSAY-----PGAPQVPAPNSFIPPGTGGFPQPTAGSFPQPGVPQVPSGSP 256
QY 92 --YVGP-----LYPGSTVLVEGGYDAGARFCAGATA--GNIPP--PPPCCPNAAQLAV 141
DB 257 ADIPQGMPPRAPPGASAPVAPGMPGAFPPQGGGGMFGSPPGAPGPGP----- 308
QY 142 MOGANVLTQRKGNFFMG-----GSDGGY 165
DB 309 -----GMPGSPAPGAPGPGPGGY 327

Search completed: March 6, 2001, 11:54:50
Job time: 247 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 11:52:12 ; Search time 34.57 Seconds
(without alignments)
188.557 Million cell updates/sec

Title: US-09-407-430-2
Perfect score: 2058
Sequence: 1 ETHVTGSGAGTGVGFVSL.....ADARVCSCLMMLLSQAEA 363

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues
Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/PCUOS_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2058	100.0	480	1	US-08-440-103-14
2	2058	100.0	480	1	US-08-440-542-14
3	2058	100.0	480	1	US-08-231-368-14
4	2058	100.0	480	1	US-08-440-210-14
5	2058	100.0	663	3	US-08-824-057-3
6	2058	100.0	2995	3	US-08-444-818-138
7	2058	100.0	3011	1	US-08-440-103-36
8	2058	100.0	3011	1	US-08-440-542-36
9	2058	100.0	3011	1	US-07-910-760-10
10	2058	100.0	3011	1	US-08-440-519-10
11	2058	100.0	3011	1	US-08-231-368-36
12	2058	100.0	3011	1	US-08-440-210-36
13	2055	99.9	2772	3	US-08-444-818-89
14	2053	99.8	2955	2	US-08-443-260-3
15	2053	99.8	2955	3	US-08-442-805A-3
16	2053	99.8	2955	3	US-08-443-900A-3
17	2053	99.8	2955	3	US-08-444-818-124
18	2053	99.8	3011	2	US-08-833-678A-6
19	2053	99.8	3011	3	US-08-444-818-177
20	2053	99.8	3011	4	PCT-US91-02225-10
21	2037	99.0	738	4	PCT-US92-06965A-5
22	2036	98.9	402	1	US-08-460-806-13
23	2036	98.9	402	1	US-08-325-630-13
24	2014	97.9	403	2	US-08-483-695-39
25	2014	97.9	403	2	US-07-965-285-39
26	2014	97.9	403	2	US-08-487-231-39
27	1982	96.3	2894	2	US-08-466-975A-23
28	1982	96.3	2894	2	US-08-391-671A-23

29	1982	96.3	2894	3	US-08-467-902A-23
30	1957	95.1	409	1	US-08-440-103-21
31	1957	95.1	409	1	US-08-440-103-24
32	1957	95.1	409	1	US-08-440-542-21
33	1957	95.1	409	1	US-08-440-542-24
34	1957	95.1	409	1	US-08-231-368-21
35	1957	95.1	409	1	US-08-231-368-24
36	1957	95.1	409	1	US-08-440-210-21
37	1957	95.1	409	1	US-08-440-210-24
38	1956	95.0	621	1	US-07-748-292-7
39	1956	95.0	622	4	PCT-US92-06965A-4
40	1955	95.0	3011	3	US-08-811-566-20
41	1952	94.8	967	1	US-08-188-281B-13
42	1952	94.8	967	4	PCT-US94-07280-13
43	1952	94.8	967	4	PCT-US95-01087-13
44	1952	94.8	1648	1	US-08-188-281B-12
45	1952	94.8	1648	4	PCT-US94-07280-12

ALIGNMENTS

RESULT 1
US-08-440-103-14
; Sequence 14, Application US/08440103
; Patent No. 5670152
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,103
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 480 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-440-103-14

Query Match 100.0%; Score 2058; DB 1; Length 480;
Best Local Similarity 100.0%; Pred. No. 2.4e-196;
Matches 363; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 1 ETHVTGSGAGHTVSGFVSLAPGAKQNVQLINTNGSWHLNSTALNCNDSLNTGWLGLFY 60
Db 15 ETHVTGSGAGHTVSGFVSLAPGAKQNVQLINTNGSWHLNSTALNCNDSLNTGWLGLFY 74
QY 61 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDPQRPYCHWHYPPKPGIIPAKSVC 120
Db 75 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDPQRPYCHWHYPPKPGIIPAKSVC 134
QY 121 GPVYCFPTSPVVVGTDRSGAPTSWGENDTDFVLNNTRPPLGNWFGCTWMNSTGFTKV 180
Db 135 GPVYCFPTSPVVVGTDRSGAPTSWGENDTDFVLNNTRPPLGNWFGCTWMNSTGFTKV 194
QY 181 CGAPPCVIGGAGNNTLHCPDTCFRKHPDATYSRCGSGPWITPRCLVDYPYRLWHYPCTIN 240
Db 195 CGAPPCVIGGAGNNTLHCPDTCFRKHPDATYSRCGSGPWITPRCLVDYPYRLWHYPCTIN 254
QY 241 YTFKIRMYGGVEHRLAECNWTGRGCDLEDRDRSELSPLLLTTTQWQVLPSCFTTLP 300
Db 255 YTFKIRMYGGVEHRLAECNWTGRGCDLEDRDRSELSPLLLTTTQWQVLPSCFTTLP 314
QY 301 ALSTGLHLHQNIVDVOYLXGVSSIASWAIKWEYVYVLLFLLADARVCSCWMLLSQ 360
Db 315 ALSTGLHLHQNIVDVOYLXGVSSIASWAIKWEYVYVLLFLLADARVCSCWMLLSQ 374
QY 361 AEA 363
Db 375 AEA 377

RESULT 2

US-08-440-542-14
; Sequence 14, Application US/08440542
; Patent No. 5670153
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,542
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 480 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-440-542-14

Query Match 100.0%; Score 2058; DB 1: Length 480;
Best Local Similarity 100.0%; Pred. No. 2.4e-196;
Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETHVTGSGAGHTVSGFVSLAPGAKQNVQLINTNGSWHLNSTALNCNDSLNTGWLGLFY 60
Db 15 ETHVTGSGAGHTVSGFVSLAPGAKQNVQLINTNGSWHLNSTALNCNDSLNTGWLGLFY 74
QY 61 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDPQRPYCHWHYPPKPGIIPAKSVC 120
Db 75 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDPQRPYCHWHYPPKPGIIPAKSVC 134
QY 121 GPVYCFPTSPVVVGTDRSGAPTSWGENDTDFVLNNTRPPLGNWFGCTWMNSTGFTKV 180
Db 135 GPVYCFPTSPVVVGTDRSGAPTSWGENDTDFVLNNTRPPLGNWFGCTWMNSTGFTKV 194
QY 181 CGAPPCVIGGAGNNTLHCPDTCFRKHPDATYSRCGSGPWITPRCLVDYPYRLWHYPCTIN 240
Db 195 CGAPPCVIGGAGNNTLHCPDTCFRKHPDATYSRCGSGPWITPRCLVDYPYRLWHYPCTIN 254
QY 241 YTFKIRMYGGVEHRLAECNWTGRGCDLEDRDRSELSPLLLTTTQWQVLPSCFTTLP 300
Db 255 YTFKIRMYGGVEHRLAECNWTGRGCDLEDRDRSELSPLLLTTTQWQVLPSCFTTLP 314
QY 301 ALSTGLHLHQNIVDVOYLXGVSSIASWAIKWEYVYVLLFLLADARVCSCWMLLSQ 360
Db 315 ALSTGLHLHQNIVDVOYLXGVSSIASWAIKWEYVYVLLFLLADARVCSCWMLLSQ 374
QY 361 AEA 363
Db 375 AEA 377

RESULT 3

US-08-231-368-14
; Sequence 14, Application US/08231368
; Patent No. 5756312
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:

LENGTH: 480 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-231-368-14

Query Match 100.0%; Score 2058; DB 1; Length 480;
Best Local Similarity 100.0%; Pred. No. 2.4e-196;
Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETHVTGSGAGHTVSGFVSLAPGAKQNVOLINTNGSMHLNSTALNCNDSLNTGWLGLFY 60
DB 15 ETHVTGSGAGHTVSGFVSLAPGAKQNVOLINTNGSMHLNSTALNCNDSLNTGWLGLFY 74

QY 61 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDPQRPYCHWYPPKPGIVPAKSVC 120
DB 75 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDPQRPYCHWYPPKPGIVPAKSVC 134

QY 121 GPVYCTPSPVVVGTTRDSGAPTSYSGENDTDVFLNTRPPLGNFPGCTWMNSTGFTKV 180
DB 135 GPVYCTPSPVVVGTTRDSGAPTSYSGENDTDVFLNTRPPLGNFPGCTWMNSTGFTKV 194

QY 181 CGAPPCVIGAGNNTLHCPTDCFRKHPDATYSRGSGPMTTPRCLVDYPRYLWHYPCTIN 240
DB 195 CGAPPCVIGAGNNTLHCPTDCFRKHPDATYSRGSGPMTTPRCLVDYPRYLWHYPCTIN 254

QY 241 YTFIKRMVYGVGVEHRLAECNTRGERCDLDRSELSPLLLTTTQOVLPSCFTTLP 300
DB 255 YTFIKRMVYGVGVEHRLAECNTRGERCDLDRSELSPLLLTTTQOVLPSCFTTLP 314

QY 301 ALSTGLIHLHQNIVDVQYLYGVGSSIASWAIRKWEYVVLFLLLADARVCSCLMMLLISQ 360
DB 315 ALSTGLIHLHQNIVDVQYLYGVGSSIASWAIRKWEYVVLFLLLADARVCSCLMMLLISQ 374

QY 361 AEA 363
DB 375 AEA 377

RESULT 4
US-08-440-210-14
Sequence 14, Application US/08440210
Patent No. 5766845
GENERAL INFORMATION:
APPLICANT: Welner, Amy J.
APPLICANT: Houghton, Michael
TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/440,210
FILING DATE: 12-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/231,368
FILING DATE:
APPLICATION NUMBER: US 07/759,575
FILING DATE: 13-SEP-1991
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.

REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0205.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 480 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-440-210-14

Query Match 100.0%; Score 2058; DB 1; Length 480;
Best Local Similarity 100.0%; Pred. No. 2.4e-196;
Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETHVTGSGAGHTVSGFVSLAPGAKQNVOLINTNGSMHLNSTALNCNDSLNTGWLGLFY 60
DB 15 ETHVTGSGAGHTVSGFVSLAPGAKQNVOLINTNGSMHLNSTALNCNDSLNTGWLGLFY 74

QY 61 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDPQRPYCHWYPPKPGIVPAKSVC 120
DB 75 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDPQRPYCHWYPPKPGIVPAKSVC 134

QY 121 GPVYCTPSPVVVGTTRDSGAPTSYSGENDTDVFLNTRPPLGNFPGCTWMNSTGFTKV 180
DB 135 GPVYCTPSPVVVGTTRDSGAPTSYSGENDTDVFLNTRPPLGNFPGCTWMNSTGFTKV 194

QY 181 CGAPPCVIGAGNNTLHCPTDCFRKHPDATYSRGSGPMTTPRCLVDYPRYLWHYPCTIN 240
DB 195 CGAPPCVIGAGNNTLHCPTDCFRKHPDATYSRGSGPMTTPRCLVDYPRYLWHYPCTIN 254

QY 241 YTFIKRMVYGVGVEHRLAECNTRGERCDLDRSELSPLLLTTTQOVLPSCFTTLP 300
DB 255 YTFIKRMVYGVGVEHRLAECNTRGERCDLDRSELSPLLLTTTQOVLPSCFTTLP 314

QY 301 ALSTGLIHLHQNIVDVQYLYGVGSSIASWAIRKWEYVVLFLLLADARVCSCLMMLLISQ 360
DB 315 ALSTGLIHLHQNIVDVQYLYGVGSSIASWAIRKWEYVVLFLLLADARVCSCLMMLLISQ 374

QY 361 AEA 363
DB 375 AEA 377

RESULT 5
US-08-824-057-3
Sequence 3, Application US/08824057
Patent No. 6121020
GENERAL INFORMATION:
APPLICANT: SELBY, MARK
APPLICANT: HOUGHTON, MICHAEL
TITLE OF INVENTION: NOVEL HEPATITIS C EI AND E2 TRUNCATED
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: CHIRON CORPORATION
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/824,057
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA: US/08/282,959
APPLICATION NUMBER: 29-JUL-1994
FILING DATE: 29-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0987.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 663 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-824-057-3

Query Match 100.0%; Score 2058; DB 3; Length 663;
Best Local Similarity 100.0%; Pred. No. 3.7e-196;
Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETHVTGSGAGHTVSGFVSLAPGAKQNVQLINTNGSWHLNSTALNCNDSLTGWLGLFY 60
DB 21 ETHVTGSGAGHTVSGFVSLAPGAKQNVQLINTNGSWHLNSTALNCNDSLTGWLGLFY 80
QY 61 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDPQPCYCHYPPKPGIIVPAKSVC 120
DB 81 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDPQPCYCHYPPKPGIIVPAKSVC 140
QY 121 GPVYCFTPSPVVGTTDRSGAPTSWGENDTDFVLNTRPLGNWFGCTWMNSTGFTKV 180
DB 141 GPVYCFTPSPVVGTTDRSGAPTSWGENDTDFVLNTRPLGNWFGCTWMNSTGFTKV 200
QY 181 CGAPPCVTGGAGNNTLHCPDTCFRKHDPATYSRCGSGPWITPRCLVDYPYRLWHYPCTIN 240
DB 201 CGAPPCVTGGAGNNTLHCPDTCFRKHDPATYSRCGSGPWITPRCLVDYPYRLWHYPCTIN 260
QY 241 YTIKIRMYGVGVEHRLAECNWTGRCDCLEDRDRSELSPLLLTTTQWQVLPSCFTTLP 300
DB 261 YTIKIRMYGVGVEHRLAECNWTGRCDCLEDRDRSELSPLLLTTTQWQVLPSCFTTLP 320
QY 301 ALSTGLIHLHONIVDQVLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLNWMMLLSQ 360
DB 321 ALSTGLIHLHONIVDQVLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLNWMMLLSQ 380
QY 361 AEA 363
DB 381 AEA 383

RESULT 6
US-08-444-818-138
Sequence 138, Application US/08444818
Patent No. 6150087
GENERAL INFORMATION:
APPLICANT: Chien, David Y.
APPLICANT: Rutter, William J.
TITLE OF INVENTION: NANV Diagnostics and Vaccines
NUMBER OF SEQUENCES: 777
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/444,818
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/403,590
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Harbin, Alisa A.
REGISTRATION NUMBER: 33,895
REFERENCE/DOCKET NUMBER: 0110.002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508) 359-3876
TELEFAX: (508) 359-3885
INFORMATION FOR SEQ ID NO: 138:
SEQUENCE CHARACTERISTICS:
LENGTH: 2995 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-444-818-138

Query Match 100.0%; Score 2058; DB 3; Length 2995;
Best Local Similarity 100.0%; Pred. No. 3.2e-195;
Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETHVTGSGAGHTVSGFVSLAPGAKQNVQLINTNGSWHLNSTALNCNDSLTGWLGLFY 60
DB 384 ETHVTGSGAGHTVSGFVSLAPGAKQNVQLINTNGSWHLNSTALNCNDSLTGWLGLFY 443
QY 61 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDPQPCYCHYPPKPGIIVPAKSVC 120
DB 444 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDPQPCYCHYPPKPGIIVPAKSVC 503
QY 121 GPVYCFTPSPVVGTTDRSGAPTSWGENDTDFVLNTRPLGNWFGCTWMNSTGFTKV 180
DB 504 GPVYCFTPSPVVGTTDRSGAPTSWGENDTDFVLNTRPLGNWFGCTWMNSTGFTKV 563
QY 181 CGAPPCVTGGAGNNTLHCPDTCFRKHDPATYSRCGSGPWITPRCLVDYPYRLWHYPCTIN 240
DB 564 CGAPPCVTGGAGNNTLHCPDTCFRKHDPATYSRCGSGPWITPRCLVDYPYRLWHYPCTIN 623
QY 241 YTIKIRMYGVGVEHRLAECNWTGRCDCLEDRDRSELSPLLLTTTQWQVLPSCFTTLP 300
DB 624 YTIKIRMYGVGVEHRLAECNWTGRCDCLEDRDRSELSPLLLTTTQWQVLPSCFTTLP 683
QY 301 ALSTGLIHLHONIVDQVLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLNWMMLLSQ 360
DB 684 ALSTGLIHLHONIVDQVLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLNWMMLLSQ 743
QY 361 AEA 363
DB 744 AEA 746

RESULT 7
US-08-440-103-36
Sequence 36, Application US/08440103
Patent No. 5670152
GENERAL INFORMATION:
APPLICANT: Weiner, Amy J.
APPLICANT: Houghton, Michael
TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA

ZIP: 94608
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/440,103
FILING DATE: 12-MAY-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/231,368
FILING DATE:
FILING DATE: 13-SEP-1991
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0205.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 3011 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-440-103-36

Query Match 100.0%; Score 2058; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 3.2e-195;
Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETHVTGSGAGHTVSGFVSLAPGAKONVOLINTNGSHLNTALNCNDSLNTGWLGLFY 60
DB 384 ETHVTGSGAGHTVSGFVSLAPGAKONVOLINTNGSHLNTALNCNDSLNTGWLGLFY 443

QY 61 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDQRPYCWHPKPGIVPAKSYVC 120
DB 444 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDQRPYCWHPKPGIVPAKSYVC 503

QY 121 GPVYCFPTSPVVVGTTRDSGAPTSYSGENDTDVFLNTRPPLGNFGCTWMNSTGFTKV 180
DB 504 GPVYCFPTSPVVVGTTRDSGAPTSYSGENDTDVFLNTRPPLGNFGCTWMNSTGFTKV 563

QY 181 CGAPPCVIGGAGNNTLHCPTDCFRKHPDATYSRCGSGPWITPRCLVDYPRYLWHYPCTIN 240
DB 564 CGAPPCVIGGAGNNTLHCPTDCFRKHPDATYSRCGSGPWITPRCLVDYPRYLWHYPCTIN 623

QY 241 YTFIKRMVYGVGVEHRELAACNWTGERCDLEDRDRSELSPLLLTTTQOVLPCTIN 300
DB 624 YTFIKRMVYGVGVEHRELAACNWTGERCDLEDRDRSELSPLLLTTTQOVLPCTIN 683

QY 301 ALSTGLIHLHQNIVDQYLYGVGSSIASNAIKWEYVLLFLLADARVCSCLMMMLLSQ 360
DB 684 ALSTGLIHLHQNIVDQYLYGVGSSIASNAIKWEYVLLFLLADARVCSCLMMMLLSQ 743

QY 361 AEA 363
DB 744 AEA 746

RESULT 8
US-08-440-542-36
Sequence 36, Application US/08440542
Patent No. 5670153
GENERAL INFORMATION:
APPLICANT: Welner, Amy J.
APPLICANT: Houghton, Michael
TITLE OF INVENTION: Immunoreactive Polypeptide Compositions

NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/440,542
FILING DATE: 12-MAY-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/231,368
FILING DATE:
APPLICATION NUMBER: US 07/759,575
FILING DATE: 13-SEP-1991
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0205.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 3011 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-440-542-36

Query Match 100.0%; Score 2058; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 3.2e-195;
Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETHVTGSGAGHTVSGFVSLAPGAKONVOLINTNGSHLNTALNCNDSLNTGWLGLFY 60
DB 384 ETHVTGSGAGHTVSGFVSLAPGAKONVOLINTNGSHLNTALNCNDSLNTGWLGLFY 443

QY 61 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDQRPYCWHPKPGIVPAKSYVC 120
DB 444 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDQRPYCWHPKPGIVPAKSYVC 503

QY 121 GPVYCFPTSPVVVGTTRDSGAPTSYSGENDTDVFLNTRPPLGNFGCTWMNSTGFTKV 180
DB 504 GPVYCFPTSPVVVGTTRDSGAPTSYSGENDTDVFLNTRPPLGNFGCTWMNSTGFTKV 563

QY 181 CGAPPCVIGGAGNNTLHCPTDCFRKHPDATYSRCGSGPWITPRCLVDYPRYLWHYPCTIN 240
DB 564 CGAPPCVIGGAGNNTLHCPTDCFRKHPDATYSRCGSGPWITPRCLVDYPRYLWHYPCTIN 623

QY 241 YTFIKRMVYGVGVEHRELAACNWTGERCDLEDRDRSELSPLLLTTTQOVLPCTIN 300
DB 624 YTFIKRMVYGVGVEHRELAACNWTGERCDLEDRDRSELSPLLLTTTQOVLPCTIN 683

QY 301 ALSTGLIHLHQNIVDQYLYGVGSSIASNAIKWEYVLLFLLADARVCSCLMMMLLSQ 360
DB 684 ALSTGLIHLHQNIVDQYLYGVGSSIASNAIKWEYVLLFLLADARVCSCLMMMLLSQ 743

QY 361 AEA 363
DB 744 AEA 746

RESULT 9

US-07-910-760-10
; Sequence 10, Application US/07910760
; Patent No. 5683864
; GENERAL INFORMATION:
; APPLICANT: Houghton, Michael
; APPLICANT: Choo, Qui-Lim
; APPLICANT: Kuo, George
; TITLE OF INVENTION: Combinations of Hepatitis C virus (HCV)
; TITLE OF INVENTION: Antigens for Use in Immunoassays for Anti-HCV Antibodies
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: P.O. Box 8097 (Int. Prop. R-440)
; CITY: Emeryville
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 94662-8097
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/07/910,760
; APPLICATION NUMBER: US/07/910,760
; FILING DATE: 07-JUL-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Blackburn Esq., Robert P.
; REGISTRATION NUMBER: 30,447
; REFERENCE/DOCKET NUMBER: 0101.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2702
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3011 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-910-760-10

Query Match 100.0%; Score 2058; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 3.2e-195;
Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ETHVTGSGAGTGVSGFVSLAPGAKQNVQLINTNGSWHLNLTALNCNDSLNTGLAGLFY 60
Db 384 ETHVTGSGAGTGVSGFVSLAPGAKQNVQLINTNGSWHLNLTALNCNDSLNTGLAGLFY 443
Qy 61 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGPQRPYCHWHYPPKPGCIIVPAKSVC 120
Db 444 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGPQRPYCHWHYPPKPGCIIVPAKSVC 503
Qy 121 GPVYCFTSPVVVGTDRSGAPTSWGENDTDFVNLNTRPPLGNWFGCTWMNSTGFTKV 180
Db 504 GPVYCFTSPVVVGTDRSGAPTSWGENDTDFVNLNTRPPLGNWFGCTWMNSTGFTKV 563
Qy 181 CGAPPCVIGGAGNNTLHCPTDCFRKHPDATYSCGSGPWIITPRCLVDYPIRLWHYPCIN 240
Db 564 CGAPPCVIGGAGNNTLHCPTDCFRKHPDATYSCGSGPWIITPRCLVDYPIRLWHYPCIN 623
Qy 241 YTIKIRMYGVGVEHRLAECNWTGERCDLEDRDRSELSPLLLTTTQWVLPSCFTTLP 300
Db 624 YTIKIRMYGVGVEHRLAECNWTGERCDLEDRDRSELSPLLLTTTQWVLPSCFTTLP 683
Qy 301 ALSTGLIHLHQNIVDQVLYGVGSSIASWAIKWEYVLLFLLADARVCSCLMMMLLSQ 360
Db 684 ALSTGLIHLHQNIVDQVLYGVGSSIASWAIKWEYVLLFLLADARVCSCLMMMLLSQ 743
Qy 361 AEA 363
Db 744 AEA 746

RESULT 10
US-08-440-519-10
; Sequence 10, Application US/08440519
; Patent No. 5712087
; GENERAL INFORMATION:
; APPLICANT: Houghton, Michael
; APPLICANT: Choo, Qui-Lim
; APPLICANT: Kuo, George
; TITLE OF INVENTION: Combinations of Hepatitis C virus (HCV)
; TITLE OF INVENTION: Antigens for Use in Immunoassays for Anti-HCV Antibodies
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: P.O. Box 8097 (Int. Prop. R-440)
; CITY: Emeryville
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 94662-8097
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/440,519
; APPLICATION NUMBER: US/08/440,519
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/910,760
; FILING DATE: 07-JUL-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Blackburn Esq., Robert P.
; REGISTRATION NUMBER: 30,447
; REFERENCE/DOCKET NUMBER: 0101.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2702
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3011 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-440-519-10

Query Match 100.0%; Score 2058; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 3.2e-195;
Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ETHVTGSGAGTGVSGFVSLAPGAKQNVQLINTNGSWHLNLTALNCNDSLNTGLAGLFY 60
Db 384 ETHVTGSGAGTGVSGFVSLAPGAKQNVQLINTNGSWHLNLTALNCNDSLNTGLAGLFY 443
Qy 61 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGPQRPYCHWHYPPKPGCIIVPAKSVC 120
Db 444 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGPQRPYCHWHYPPKPGCIIVPAKSVC 503
Qy 121 GPVYCFTSPVVVGTDRSGAPTSWGENDTDFVNLNTRPPLGNWFGCTWMNSTGFTKV 180
Db 504 GPVYCFTSPVVVGTDRSGAPTSWGENDTDFVNLNTRPPLGNWFGCTWMNSTGFTKV 563
Qy 181 CGAPPCVIGGAGNNTLHCPTDCFRKHPDATYSCGSGPWIITPRCLVDYPIRLWHYPCIN 240
Db 564 CGAPPCVIGGAGNNTLHCPTDCFRKHPDATYSCGSGPWIITPRCLVDYPIRLWHYPCIN 623
Qy 241 YTIKIRMYGVGVEHRLAECNWTGERCDLEDRDRSELSPLLLTTTQWVLPSCFTTLP 300
Db 624 YTIKIRMYGVGVEHRLAECNWTGERCDLEDRDRSELSPLLLTTTQWVLPSCFTTLP 683
Qy 301 ALSTGLIHLHQNIVDQVLYGVGSSIASWAIKWEYVLLFLLADARVCSCLMMMLLSQ 360

Db 684 ALSTGLIHLHONIVDVQYLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLWMLLSQ 743
QY 361 AEA 363
Db 744 AEA 746
RESULT 11
US-08-231-368-36
: Sequence 36, Application US/08231368
: Patent No. 5756312
: GENERAL INFORMATION:
: APPLICANT: Weiner, Amy J.
: APPLICANT: Houghton, Michael
: TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
: NUMBER OF SEQUENCES: 45
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Chiron Corporation
: STREET: 4560 Horton Street
: CITY: Emeryville
: STATE: CA
: COUNTRY: USA
: ZIP: 94608
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/231,368
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/759,575
: FILING DATE: 13-SEP-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: McClung, Barbara G.
: REGISTRATION NUMBER: 33,113
: REFERENCE/DOCKET NUMBER: 0205.001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (510) 601-2708
: TELEFAX: (510) 655-3542
: INFORMATION FOR SEQ ID NO: 36:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3011 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-231-368-36
Query Match 100.0%; Score 2058; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 3.2e-195;
Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ETHVTGGSAGHTVSGFVSLAPGAKONVOLINTNGSWHLNSTALNCNDSLNTGWLGLF 60
Db 384 ETHVTGGSAGHTVSGFVSLAPGAKONVOLINTNGSWHLNSTALNCNDSLNTGWLGLF 443
QY 61 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDQRPYCHWYPPKPGIIVPAKSV 120
Db 444 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDQRPYCHWYPPKPGIIVPAKSV 503
QY 121 GPVYCFTSPVVVGTDRSGAPTSYSGENDTDVFLNNTRPPLGNWFGCTWNNSTGFTKV 180
Db 504 GPVYCFTSPVVVGTDRSGAPTSYSGENDTDVFLNNTRPPLGNWFGCTWNNSTGFTKV 563
QY 181 CGAPPCVIGGAGNNTLHCPTDCFRKHPDQATYSCGSGPWITPRCLVDYPIRLWHYPCPTIN 240
Db 564 CGAPPCVIGGAGNNTLHCPTDCFRKHPDQATYSCGSGPWITPRCLVDYPIRLWHYPCPTIN 623

QY 241 YTIKIRMYGGVEHRLLEAAACNNTRGRCDEDRDRSELSPLLLTTTQWQVLPCTFTLP 300
Db 624 YTIKIRMYGGVEHRLLEAAACNNTRGRCDEDRDRSELSPLLLTTTQWQVLPCTFTLP 683
QY 301 ALSTGLIHLHONIVDVQYLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLWMLLSQ 360
Db 684 ALSTGLIHLHONIVDVQYLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLWMLLSQ 743
QY 361 AEA 363
Db 744 AEA 746
RESULT 12
US-08-440-210-36
: Sequence 36, Application US/08440210
: Patent No. 5766845
: GENERAL INFORMATION:
: APPLICANT: Weiner, Amy J.
: APPLICANT: Houghton, Michael
: TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
: NUMBER OF SEQUENCES: 45
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Chiron Corporation
: STREET: 4560 Horton Street
: CITY: Emeryville
: STATE: CA
: COUNTRY: USA
: ZIP: 94608
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/440,210
: FILING DATE: 12-MAY-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/231,368
: FILING DATE:
: APPLICATION NUMBER: US 07/759,575
: FILING DATE: 13-SEP-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: McClung, Barbara G.
: REGISTRATION NUMBER: 33,113
: REFERENCE/DOCKET NUMBER: 0205.001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (510) 601-2708
: TELEFAX: (510) 655-3542
: INFORMATION FOR SEQ ID NO: 36:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3011 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-440-210-36
Query Match 100.0%; Score 2058; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 3.2e-195;
Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ETHVTGGSAGHTVSGFVSLAPGAKONVOLINTNGSWHLNSTALNCNDSLNTGWLGLF 60
Db 384 ETHVTGGSAGHTVSGFVSLAPGAKONVOLINTNGSWHLNSTALNCNDSLNTGWLGLF 443
QY 61 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDQRPYCHWYPPKPGIIVPAKSV 120
Db 444 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDQRPYCHWYPPKPGIIVPAKSV 503
QY 121 GPVYCFTSPVVVGTDRSGAPTSYSGENDTDVFLNNTRPPLGNWFGCTWNNSTGFTKV 180

Db 504 GPVYCFPSVVVGTDRSGAPYSWGENDTDFVNLNTRPPLGNWFGCTWMNSTGFTKV 563
QY 181 CGAPPCVIGGAGNNTLHCPDTCFRKHPDATYSRGSGFWITPRCLVDYPRYRLWHYPCTIN 240
Db 564 CGAPPCVIGGAGNNTLHCPDTCFRKHPDATYSRGSGFWITPRCLVDYPRYRLWHYPCTIN 623
QY 241 YTIKIRMYGGVHRLEAACNWTGRGCDLEDRSELSPLLLTTTQWQVLPSCFTTLP 300
Db 624 YTIKIRMYGGVHRLEAACNWTGRGCDLEDRSELSPLLLTTTQWQVLPSCFTTLP 683
QY 301 ALSTGLHLHQNIVDQYLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLMMMLLSQ 360
Db 684 ALSTGLHLHQNIVDQYLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLMMMLLSQ 743
QY 361 AEA 363
Db 744 AEA 746

RESULT 13
US-08-444-818-89
; Sequence 89, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
; APPLICANT: Chien, David Y.
; APPLICANT: Rutter, William J.
; TITLE OF INVENTION: NANBV Diagnostics and Vaccines
; NUMBER OF SEQUENCES: 777
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444.818
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/403.590
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0110.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508)359-3876
; TELEFAX: (508)359-3885
; INFORMATION FOR SEQ ID NO: 89:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2772 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-444-818-89

Query Match 99.9%; Score 2055; DB 3; Length 2772;
Best Local Similarity 99.7%; Pred. No. 5.7e-195;
Matches 362; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 ETHVTGSGAGHTVSGFVSLAPAKQNVQLINTNGSWHLNLTALNCNDSLNTGWLGLFY 60
Db 270 ETHVTGSGAGHTVSGFVSLAPAKQNVQLINTNGSWHLNLTALNCNDSLNTGWLGLFY 329
QY 61 HHKFNSSGCPERLASCRPLTDFQGWGPISYANGSGPDORPYCWHYPKPCGIVPAKSVC 120

Db 330 HHKFNSSGCPERLASCRPLTDFQGWGPISYANGSGPDORPYCWHYPKPCGIVPAKSVC 389
QY 121 GPVYCFPSVVVGTDRSGAPYSWGENDTDFVNLNTRPPLGNWFGCTWMNSTGFTKV 180
Db 390 GPVYCFPSVVVGTDRSGAPYSWGENDTDFVNLNTRPPLGNWFGCTWMNSTGFTKV 449
QY 181 CGAPPCVIGGAGNNTLHCPDTCFRKHPDATYSRGSGFWITPRCLVDYPRYRLWHYPCTIN 240
Db 450 CGAPPCVIGGAGNNTLHCPDTCFRKHPDATYSRGSGFWITPRCLVDYPRYRLWHYPCTIN 509
QY 241 YTIKIRMYGGVHRLEAACNWTGRGCDLEDRSELSPLLLTTTQWQVLPSCFTTLP 300
Db 510 YTIKIRMYGGVHRLEAACNWTGRGCDLEDRSELSPLLLTTTQWQVLPSCFTTLP 569
QY 301 ALSTGLHLHQNIVDQYLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLMMMLLSQ 360
Db 570 ALSTGLHLHQNIVDQYLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLMMMLLSQ 629
QY 361 AEA 363
Db 630 AEA 632

RESULT 14
US-08-443-260-3
; Sequence 3, Application US/08443260
; Patent No. 5942234
; GENERAL INFORMATION:
; APPLICANT: RALSTON, ROBERT O.
; APPLICANT: MARCUS, FRANK
; APPLICANT: THUDIGM, KENT B.
; APPLICANT: GERVAISE, BARBARA A.
; APPLICANT: HALL, JOHN A.
; TITLE OF INVENTION: HEPATITIS C VIRUS ASIALOGLYCOPROTEINS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/443.260
; FILING DATE: 17-MAY-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: HARBIN, ALISA A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0154.006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2955 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 9
; OTHER INFORMATION: /note= "There is a heterogeneity at
; OTHER INFORMATION: this location; Xaa = Arg or Lys"
; FEATURE:
; NAME/KEY: Modified-site

LOCATION: 11
OTHER INFORMATION: /note= "There is a heterogeneity at
this location; Xaa = Asn or Thr"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 176
OTHER INFORMATION: /note= "There is a heterogeneity at
this location; Xaa = Ile or Thr"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 334
OTHER INFORMATION: /note= "There is a heterogeneity at
this location; Xaa = Met or Val"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 603
OTHER INFORMATION: /note= "There is a heterogeneity at
this location; Xaa = Ile or Leu"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 848
OTHER INFORMATION: /note= "There is a heterogeneity at
this location; Xaa = Asn or Tyr"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1114
OTHER INFORMATION: /note= "There is a heterogeneity at
this location; Xaa = Pro or Ser"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1117
OTHER INFORMATION: /note= "There is a heterogeneity at
this location; Xaa = Ser or Thr"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1276
OTHER INFORMATION: /note= "There is a heterogeneity at
this location; Xaa = Leu or Pro"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1454
OTHER INFORMATION: /note= "There is a heterogeneity at
this location; Xaa = Cys or Tyr"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1471
OTHER INFORMATION: /note= "There is a heterogeneity at
this location; Xaa = Ser or Thr"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1877
OTHER INFORMATION: /note= "There is a heterogeneity at
this location; Xaa = Glu or Gly"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1948
OTHER INFORMATION: /note= "There is a heterogeneity at
this location; Xaa = His or Leu"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1949
OTHER INFORMATION: /note= "There is a heterogeneity at
this location; Xaa = Cys or Ser"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 2021
OTHER INFORMATION: /note= "There is a heterogeneity at
this location; Xaa = Gly or Val"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 2349
OTHER INFORMATION: /note= "There is a heterogeneity at
this location; Xaa = Ser or Thr"

FEATURE:
NAME/KEY: Modified-site
LOCATION: 2385
OTHER INFORMATION: /note= "There is a heterogeneity at
this location; Xaa = Phe or Tyr"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 2386
OTHER INFORMATION: /note= "There is a heterogeneity at
this location; Xaa = Ala or Ser"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 2502
OTHER INFORMATION: /note= "There is a heterogeneity at
this location; Xaa = Phe or Leu"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 2690
OTHER INFORMATION: /note= "There is a heterogeneity at
this location; Xaa = Gly or Arg"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 2921
OTHER INFORMATION: /note= "There is a heterogeneity at
this location; Xaa = Arg or Gly"
US-08-443-260-3

Query Match 99.8%; Score 2053; DB 2; Length 2955;

Best Local Similarity 99.7%; Pred. No. 9.8e-195;

Matches 362; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ETHVTGSGAGHTVSGFVSLAPGAKONVOLINTNGSWHLNSTALNCNDSLNTGWLGLFY 60
DB 384 ETHVTGSGAGHTVSGFVSLAPGAKONVOLINTNGSWHLNSTALNCNDSLNTGWLGLFY 443
QY 61 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDOPRYCWHYPKPGIVPAKSYVC 120
DB 444 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDOPRYCWHYPKPGIVPAKSYVC 503
QY 121 GPVYCFTPSPVVVGTTRDSGAPTSYMGENDTDVFLNNTRPPLGNWFGCTWMNSTGFTKV 180
DB 504 GPVYCFTPSPVVVGTTRDSGAPTSYMGENDTDVFLNNTRPPLGNWFGCTWMNSTGFTKV 563
QY 181 CGAPCVIGGAGNNTLHCPTDFRKHDPDATYSRGSGPWITPRCLVDYDPRLWHYPCTIN 240
DB 564 CGAPCVIGGAGNNTLHCPTDFRKHDPDATYSRGSGPWITPRCLVDYDPRLWHYPCTIN 623
QY 241 YTIKIRMYVGGVEHRLAECNWTGERCDLDRORSELSPLLLTTTQOVLPSCFTTLP 300
DB 624 YTIKIRMYVGGVEHRLAECNWTGERCDLDRORSELSPLLLTTTQOVLPSCFTTLP 683
QY 301 ALSTGLIHLHQINVDVQVLYGVGSSIASWAIKWEYVVLLFLLADARVCSCLMMMLLSQ 360
DB 684 ALSTGLIHLHQINVDVQVLYGVGSSIASWAIKWEYVVLLFLLADARVCSCLMMMLLSQ 743
QY 361 AEA 363
DB 744 AEA 746

RESULT 15

US-08-442-805A-3

; Sequence 3, Application US/08442805A

; Patent No. 6074846

; GENERAL INFORMATION:

; APPLICANT: RALSTON, ROBERT O.

; APPLICANT: MARCUS, FRANK

; APPLICANT: THUODUM, KENT B.

; APPLICANT: GERVASE, BARBARA A.

; APPLICANT: HALL, JOHN A.

; TITLE OF INVENTION: HEPATITIS C VIRUS ASIALOGLYCOPROTEINS

; NUMBER OF SEQUENCES: 3

;;
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Chiron Corporation
;; STREET: 4560 Horton Street
;; CITY: Emeryville
;; STATE: California
;; COUNTRY: USA
;; ZIP: 94608-2916
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/442,805A
;; FILING DATE: 17-MAY-1995
;; CLASSIFICATION: 530
;; ATTORNEY/AGENT INFORMATION:
;; NAME: HARBIN, ALISA A.
;; REGISTRATION NUMBER: 33,895
;; REFERENCE/DOCKET NUMBER: 0154,005
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (510) 601-2708
;; TELEFAX: (510) 655-3542
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2955 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; FEATURE:
;; NAME/KEY: Modified-site
;; LOCATION: 9
;; OTHER INFORMATION: /note= "There is a heterogeneity at
;; OTHER INFORMATION: this location; Xaa = Arg or Lys"
;; FEATURE:
;; NAME/KEY: Modified-site
;; LOCATION: 11
;; OTHER INFORMATION: /note= "There is a heterogeneity at
;; OTHER INFORMATION: this location; Xaa = Asn or Thr"
;; FEATURE:
;; NAME/KEY: Modified-site
;; LOCATION: 176
;; OTHER INFORMATION: /note= "There is a heterogeneity at
;; OTHER INFORMATION: this location; Xaa = Ile or Thr"
;; FEATURE:
;; NAME/KEY: Modified-site
;; LOCATION: 334
;; OTHER INFORMATION: /note= "There is a heterogeneity at
;; OTHER INFORMATION: this location; Xaa = Met or Val"
;; FEATURE:
;; NAME/KEY: Modified-site
;; LOCATION: 603
;; OTHER INFORMATION: /note= "There is a heterogeneity at
;; OTHER INFORMATION: this location; Xaa = Ile or Leu"
;; FEATURE:
;; NAME/KEY: Modified-site
;; LOCATION: 848
;; OTHER INFORMATION: /note= "There is a heterogeneity at
;; OTHER INFORMATION: this location; Xaa = Asn or Tyr"
;; FEATURE:
;; NAME/KEY: Modified-site
;; LOCATION: 1114
;; OTHER INFORMATION: /note= "There is a heterogeneity at
;; OTHER INFORMATION: this location; Xaa = Pro or Ser"
;; FEATURE:
;; NAME/KEY: Modified-site
;; LOCATION: 1117
;; OTHER INFORMATION: /note= "There is a heterogeneity at
;; OTHER INFORMATION: this location; Xaa = Ser or Thr"
;; FEATURE:
;; NAME/KEY: Modified-site
;; LOCATION: 1276

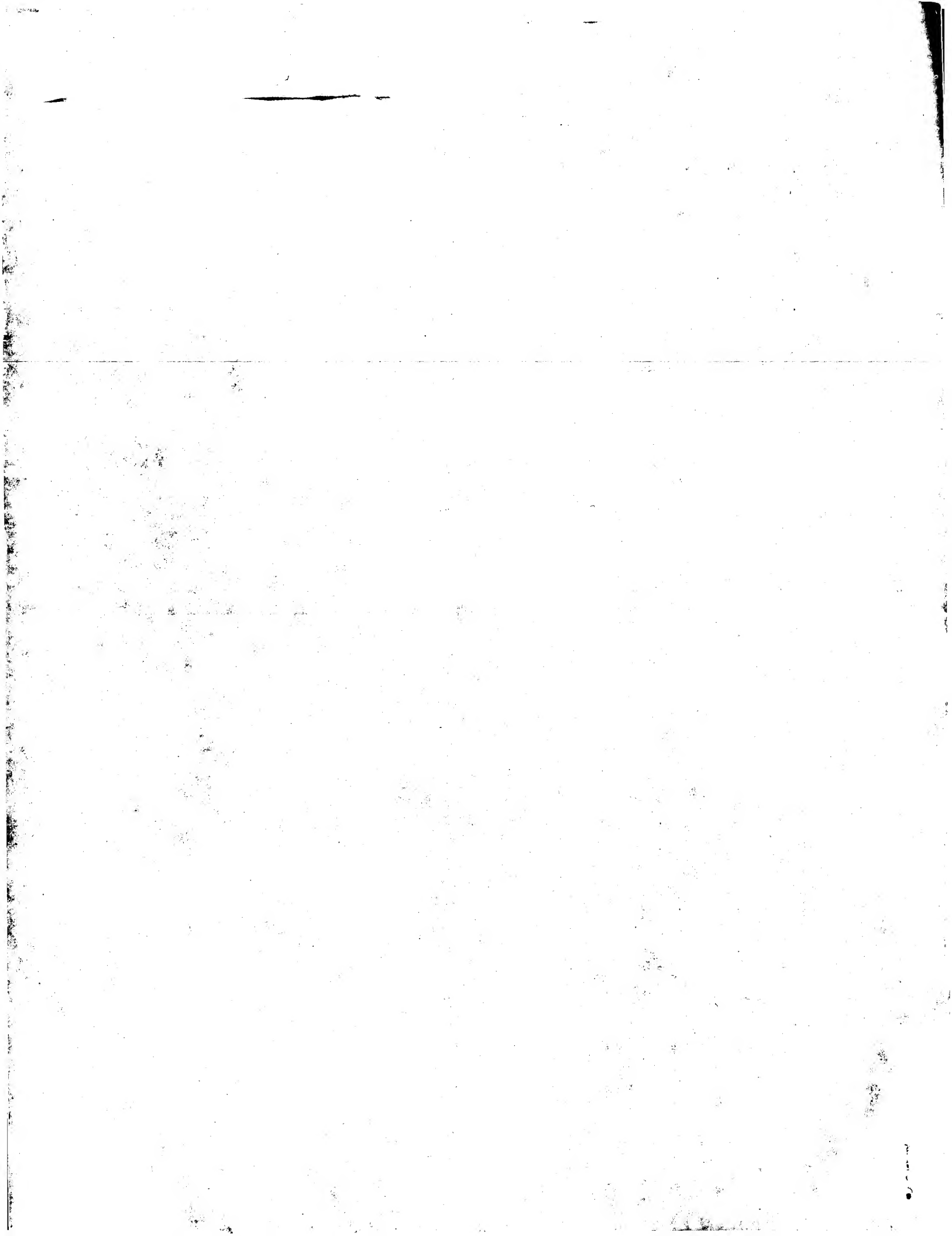
;;
;; OTHER INFORMATION: /note= "There is a heterogeneity at
;; OTHER INFORMATION: this location; Xaa = Leu or Pro"
;; FEATURE:
;; NAME/KEY: Modified-site
;; LOCATION: 1454
;; OTHER INFORMATION: /note= "There is a heterogeneity at
;; OTHER INFORMATION: this location; Xaa = Cys or Tyr"
;; FEATURE:
;; NAME/KEY: Modified-site
;; LOCATION: 1471
;; OTHER INFORMATION: /note= "There is a heterogeneity at
;; OTHER INFORMATION: this location; Xaa = Ser or Thr"
;; FEATURE:
;; NAME/KEY: Modified-site
;; LOCATION: 1877
;; OTHER INFORMATION: /note= "There is a heterogeneity at
;; OTHER INFORMATION: this location; Xaa = Glu or Gly"
;; FEATURE:
;; NAME/KEY: Modified-site
;; LOCATION: 1948
;; OTHER INFORMATION: /note= "There is a heterogeneity at
;; OTHER INFORMATION: this location; Xaa = His or Leu"
;; FEATURE:
;; NAME/KEY: Modified-site
;; LOCATION: 1949
;; OTHER INFORMATION: /note= "There is a heterogeneity at
;; OTHER INFORMATION: this location; Xaa = Cys or Ser"
;; FEATURE:
;; NAME/KEY: Modified-site
;; LOCATION: 2021
;; OTHER INFORMATION: /note= "There is a heterogeneity at
;; OTHER INFORMATION: this location; Xaa = Gly or Val"
;; FEATURE:
;; NAME/KEY: Modified-site
;; LOCATION: 2349
;; OTHER INFORMATION: /note= "There is a heterogeneity at
;; OTHER INFORMATION: this location; Xaa = Ser or Thr"
;; FEATURE:
;; NAME/KEY: Modified-site
;; LOCATION: 2385
;; OTHER INFORMATION: /note= "There is a heterogeneity at
;; OTHER INFORMATION: this location; Xaa = Phe or Tyr"
;; FEATURE:
;; NAME/KEY: Modified-site
;; LOCATION: 2386
;; OTHER INFORMATION: /note= "There is a heterogeneity at
;; OTHER INFORMATION: this location; Xaa = Ala or Ser"
;; FEATURE:
;; NAME/KEY: Modified-site
;; LOCATION: 2502
;; OTHER INFORMATION: /note= "There is a heterogeneity at
;; OTHER INFORMATION: this location; Xaa = Phe or Leu"
;; FEATURE:
;; NAME/KEY: Modified-site
;; LOCATION: 2690
;; OTHER INFORMATION: /note= "There is a heterogeneity at
;; OTHER INFORMATION: this location; Xaa = Gly or Arg"
;; FEATURE:
;; NAME/KEY: Modified-site
;; LOCATION: 2921
;; OTHER INFORMATION: /note= "There is a heterogeneity at
;; OTHER INFORMATION: this location; Xaa = Arg or Gly"
;;
;; US-08-442-805A-3

Query Match 99.8%; Score 2053; DB 3; Length 2955;
Best Local Similarity 99.7%; Pred. No. 9.8e-195;
Matches 362; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 ETHVTGSGAGHTVSGFVSLAPGAKQNVQLINTNGSWHLNLTALNCNDSLTGWLGLFY 60
DB 384 ETHVTGSGAGHTVSGFVSLAPGAKQNVQLINTNGSWHLNLTALNCNDSLTGWLGLFY 443

QY	61	HHKFNSSGCCPERLASCRPLTDFDQGWGPISYANGSGPDQRPYCWHPKPGIYPAKSVC	120
Db	444	HHKFNSSGCCPERLASCRPLTDFDQGWGPISYANGSGPDQRPYCWHPKPGIYPAKSVC	503
QY	121	GPVYCFTTSPVVVGTDRSGAPTYSWGENDTDFVNLNTRPPLGNWFGCTWMNSTGFTKV	180
Db	504	GPVYCFTTSPVVVGTDRSGAPTYSWGENDTDFVNLNTRPPLGNWFGCTWMNSTGFTKV	563
QY	181	CGAPPCVIGGAGNNTLHCPTDCFRKHDPDATYSRCGSGPWITPRCLVDYPYRLWHYPCIN	240
Db	564	CGAPPCVIGGAGNNTLHCPTDCFRKHDPDATYSRCGSGPWITPRCLVDYPYRLWHYPCIN	623
QY	241	YTIFKIRMYVGGVEHRLAAACNWTGERCDLEDRSELSPLLLTTTQWQVLPCESTTLP	300
Db	624	YTIFKIRMYVGGVEHRLAAACNWTGERCDLEDRSELSPLLLTTTQWQVLPCESTTLP	683
QY	301	ALSTGLIHLHQNIVDVQYLYGVGSSIASWAIKWEYVYVLLFLLLADARVCSCLWMLLIQ	360
Db	684	ALSTGLIHLHQNIVDVQYLYGVGSSIASWAIKWEYVYVLLFLLLADARVCSCLWMLLIQ	743
QY	361	AEA 363	
Db	744	AEA 746	

Search completed: March 6, 2001, 11:52:23
Job time: 101 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 11:52:23 ; Search time 34.57 Seconds
(without alignments)
131.938 Million cell updates/sec

Title: US-09-407-430-3
Perfect score: 1494
Sequence: 1 GAKQNVQLINTNGSWHLNST.....LEAACNWTGRGCDLEDNRDR 254

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_AA.*

- 1: /cgn2_5/ptodata/2/iaa/5A_COMB.pep.*
- 2: /cgn2_5/ptodata/2/iaa/5B_COMB.pep.*
- 3: /cgn2_5/ptodata/2/iaa/6_COMB.pep.*
- 4: /cgn2_5/ptodata/2/iaa/PCTUS_COMB.pep.*
- 5: /cgn2_5/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1494	100.0	402	1	US-08-460-806-13
2	1494	100.0	402	1	US-08-325-630-13
3	1494	100.0	403	2	US-08-483-693-39
4	1494	100.0	403	2	US-07-965-283-39
5	1494	100.0	403	2	US-08-487-231-39
6	1494	100.0	480	1	US-08-440-103-14
7	1494	100.0	480	1	US-08-440-542-14
8	1494	100.0	480	1	US-08-231-368-14
9	1494	100.0	480	1	US-08-440-210-14
10	1494	100.0	653	3	US-08-824-057-3
11	1494	100.0	2772	3	US-08-444-818-89
12	1494	100.0	2995	3	US-08-444-818-138
13	1494	100.0	3011	1	US-08-440-103-36
14	1494	100.0	3011	1	US-08-440-542-36
15	1494	100.0	3011	1	US-07-910-760-10
16	1494	100.0	3011	1	US-08-440-519-10
17	1494	100.0	3011	1	US-08-231-368-36
18	1494	100.0	3011	1	US-08-440-210-36
19	1489	99.7	2955	2	US-08-443-260-3
20	1489	99.7	2955	3	US-08-443-805A-3
21	1489	99.7	2955	3	US-08-443-900A-3
22	1489	99.7	2955	3	US-08-444-818-124
23	1489	99.7	3011	2	US-08-833-678A-6
24	1489	99.7	3011	3	US-08-444-818-177
25	1489	99.7	3011	4	PCT-US91-02225-10
26	1473	98.6	621	1	US-07-748-292-7
27	1473	98.6	622	4	PCT-US92-06965A-4
28	1473	98.6	738	4	PCT-US92-06965A-5

29	1473	98.6	2894	2	US-08-466-975A-23	Sequence 23, Appl
30	1473	98.6	2894	2	US-08-391-671A-23	Sequence 23, Appl
31	1473	98.6	2894	3	US-08-467-902A-23	Sequence 23, Appl
32	1426	95.4	409	1	US-08-440-103-21	Sequence 21, Appl
33	1426	95.4	409	1	US-08-440-103-24	Sequence 24, Appl
34	1426	95.4	409	1	US-08-440-542-21	Sequence 21, Appl
35	1426	95.4	409	1	US-08-440-542-24	Sequence 24, Appl
36	1426	95.4	409	1	US-08-231-368-21	Sequence 21, Appl
37	1426	95.4	409	1	US-08-231-368-24	Sequence 24, Appl
38	1426	95.4	409	1	US-08-440-210-21	Sequence 21, Appl
39	1426	95.4	409	1	US-08-440-210-24	Sequence 21, Appl
40	1418	94.9	3011	3	US-09-014-416-1	Sequence 1, Appl
41	1418	94.9	3012	3	US-08-811-566-2	Sequence 2, Appl
42	1416	94.8	3011	3	US-08-811-566-20	Sequence 20, Appl
43	1414	94.6	402	1	US-08-460-806-17	Sequence 17, Appl
44	1414	94.6	402	1	US-08-325-630-17	Sequence 17, Appl
45	1413	94.6	305	3	US-08-478-073-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-460-806-13
; Sequence 13, Application US/08460806
; Patent No. 5747241
; GENERAL INFORMATION:
; APPLICANT: MIYAMURA, TATSUO
; APPLICANT: SAITO, IZUMU
; APPLICANT: HARADA, SHIZUKO
; APPLICANT: HONDA, YOSHIKAZU
; TITLE OF INVENTION: DIAGNOSTIC REAGENT FOR HEPATITIS C
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT,
; ADDRESS: P.C.
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08460,806
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/325,630
; FILING DATE: 19-OCT-1994
; APPLICATION NUMBER: US 07/956,993
; FILING DATE: 06-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, No. 5747241man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 4667-001-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 402 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-460-806-13

Query Match 100.0% Score 1494; DB 1; Length 402;

Best Local Similarity 100.0%; Pred. No. 4e-141;
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAKONVOLINTNGSWHLNSTALNCNDSLTGWLGLAGLFYHHKFNSSGCGPERLASCRPLTDF 60
Db 67 GAKONVOLINTNGSWHLNSTALNCNDSLTGWLGLAGLFYHHKFNSSGCGPERLASCRPLTDF 126

QY 61 DOGWGPISYANGSGDPORPCYCHWYPPKPGCIVPAKSVCGPVYCFTPSPVVVGTDRSGAP 120
Db 127 DOGWGPISYANGSGDPORPCYCHWYPPKPGCIVPAKSVCGPVYCFTPSPVVVGTDRSGAP 186

QY 121 TYSWGENDTDFVLLNTRPPLGNWFGCTWMNSTGFTKVCGAPPCVIGGAGNNTLHCPTDC 180
Db 187 TYSWGENDTDFVLLNTRPPLGNWFGCTWMNSTGFTKVCGAPPCVIGGAGNNTLHCPTDC 246

QY 181 FRKHPDATYSRCGSGPWITPRCLVDYPRYLWHYPCTINVTIFKIRMYVGVGVEHRLAACHN 240
Db 247 FRKHPDATYSRCGSGPWITPRCLVDYPRYLWHYPCTINVTIFKIRMYVGVGVEHRLAACHN 306

QY 241 WTRGERCDLEDRDR 254
Db 307 WTRGERCDLEDRDR 320

RESULT 3
US-08-483-695-39
; Sequence 39, Application US/08483695
; Patent No. 5866139
; GENERAL INFORMATION:
; APPLICANT: Brechot, Christian
; APPLICANT: Kremsdorf, Dina
; APPLICANT: Porchon, Colette
; TITLE OF INVENTION: Nucleotide and Peptide Sequences of a
; TITLE OF INVENTION: Hepatitis C Virus Isolate, Diagnostic and Therapeutic
; TITLE OF INVENTION: Applications
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,695
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/965,285
; FILING DATE: 18-MAR-1993
; APPLICATION NUMBER: FR 91 06 882
; FILING DATE: 06-JUN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 05286-0001-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 403 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-483-695-39

Best Local Similarity 100.0%; Pred. No. 4e-141;
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAKONVOLINTNGSWHLNSTALNCNDSLTGWLGLAGLFYHHKFNSSGCGPERLASCRPLTDF 60
Db 67 GAKONVOLINTNGSWHLNSTALNCNDSLTGWLGLAGLFYHHKFNSSGCGPERLASCRPLTDF 126

QY 61 DOGWGPISYANGSGDPORPCYCHWYPPKPGCIVPAKSVCGPVYCFTPSPVVVGTDRSGAP 120
Db 127 DOGWGPISYANGSGDPORPCYCHWYPPKPGCIVPAKSVCGPVYCFTPSPVVVGTDRSGAP 186

QY 121 TYSWGENDTDFVLLNTRPPLGNWFGCTWMNSTGFTKVCGAPPCVIGGAGNNTLHCPTDC 180
Db 187 TYSWGENDTDFVLLNTRPPLGNWFGCTWMNSTGFTKVCGAPPCVIGGAGNNTLHCPTDC 246

QY 181 FRKHPDATYSRCGSGPWITPRCLVDYPRYLWHYPCTINVTIFKIRMYVGVGVEHRLAACHN 240
Db 247 FRKHPDATYSRCGSGPWITPRCLVDYPRYLWHYPCTINVTIFKIRMYVGVGVEHRLAACHN 306

QY 241 WTRGERCDLEDRDR 254
Db 307 WTRGERCDLEDRDR 320

RESULT 2
US-08-325-630-13
; Sequence 13, Application US/08325630
; Patent No. 5750331
; GENERAL INFORMATION:
; APPLICANT: MIYAMURA, TATSUO
; APPLICANT: SAITO, IZUMU
; APPLICANT: HARADA, SHIZUKO
; APPLICANT: HONDA, YOSHIKAZU
; TITLE OF INVENTION: DIAGNOSTIC REAGENT FOR HEPATITIS C
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C. Jefferson Davis Highway, Suite 400
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/325,630
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/956,993
; FILING DATE: 06-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, No. 5750331man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 4667-001-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 402 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-325-630-13

Query Match 100.0%; Score 1494; DB 1; Length 402;

Query Match 100.0%; Score 1494; DB 2; Length 403;
Best Local Similarity 100.0%; Pred. No. 4e-141;
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAKONVOLINTNGSHLNTALNCNDSLNTGWLGLFYHHKFNSSGCGPERLASCRPLTDF 60
DB 73 GAKONVOLINTNGSHLNTALNCNDSLNTGWLGLFYHHKFNSSGCGPERLASCRPLTDF 132

QY 61 DOGWGPISYANGSGDPORPCWHPKPCGIVPAKSVCGPVYCFPTSPVVVGTDRSGAP 120
DB 133 DOGWGPISYANGSGDPORPCWHPKPCGIVPAKSVCGPVYCFPTSPVVVGTDRSGAP 192

QY 121 TYSWGENDTDFVLNNTRPPLGNWFGCTWMNSTGFTKVCAGAPPCVIGGAGNNTLHCPTDC 180
DB 193 TYSWGENDTDFVLNNTRPPLGNWFGCTWMNSTGFTKVCAGAPPCVIGGAGNNTLHCPTDC 252

QY 181 FRKHDPATYSRCGSGPWITPRCLVDYPYRLWHYPCTINTYFIKIRMYVGGVEHRLAECN 240
DB 253 FRKHDPATYSRCGSGPWITPRCLVDYPYRLWHYPCTINTYFIKIRMYVGGVEHRLAECN 312

QY 241 WTRGERCDLEDRDR 254
DB 313 WTRGERCDLEDRDR 326

RESULT 4
US-07-965-285-39
; Sequence 39, Application US/07965285
; Patent No. 5879904
; GENERAL INFORMATION:
; APPLICANT: Brechot, Christian
; APPLICANT: Kremsdorf, Dina
; APPLICANT: Porchon, Colette
; TITLE OF INVENTION: Nucleotide and Peptide Sequences of a
; TITLE OF INVENTION: Hepatitis C Virus Isolate, Diagnostic and Therapeutic
; TITLE OF INVENTION: Applications
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/965,285
; FILING DATE: 18-MAR-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 91 06 882
; FILING DATE: 06-JUN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 403 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-07-965-285-39

Query Match 100.0%; Score 1494; DB 2; Length 403;
Best Local Similarity 100.0%; Pred. No. 4e-141;
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAKONVOLINTNGSHLNTALNCNDSLNTGWLGLFYHHKFNSSGCGPERLASCRPLTDF 60
DB 73 GAKONVOLINTNGSHLNTALNCNDSLNTGWLGLFYHHKFNSSGCGPERLASCRPLTDF 132

QY 61 DOGWGPISYANGSGDPORPCWHPKPCGIVPAKSVCGPVYCFPTSPVVVGTDRSGAP 120
DB 133 DOGWGPISYANGSGDPORPCWHPKPCGIVPAKSVCGPVYCFPTSPVVVGTDRSGAP 192

QY 121 TYSWGENDTDFVLNNTRPPLGNWFGCTWMNSTGFTKVCAGAPPCVIGGAGNNTLHCPTDC 180
DB 193 TYSWGENDTDFVLNNTRPPLGNWFGCTWMNSTGFTKVCAGAPPCVIGGAGNNTLHCPTDC 252

QY 181 FRKHDPATYSRCGSGPWITPRCLVDYPYRLWHYPCTINTYFIKIRMYVGGVEHRLAECN 240
DB 253 FRKHDPATYSRCGSGPWITPRCLVDYPYRLWHYPCTINTYFIKIRMYVGGVEHRLAECN 312

QY 241 WTRGERCDLEDRDR 254
DB 313 WTRGERCDLEDRDR 326

RESULT 5
US-08-487-231-39
; Sequence 39, Application US/08487231
; Patent No. 5919454
; GENERAL INFORMATION:
; APPLICANT: Brechot, Christian
; APPLICANT: Kremsdorf, Dina
; APPLICANT: Porchon, Colette
; TITLE OF INVENTION: Nucleotide and Peptide Sequences of a
; TITLE OF INVENTION: Hepatitis C Virus Isolate, Diagnostic and Therapeutic
; TITLE OF INVENTION: Applications
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,231
; FILING DATE: 07-JUNE-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/965,285
; FILING DATE: 18-MAR-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 91 06 882
; FILING DATE: 06-JUN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 05286-0001-02000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 403 amino acids
; TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-487-231-39

Query Match 100.0%; Score 1494; DB 2; Length 403;
Best Local Similarity 100.0%; Pred. NO. 4e-141;
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAKNQVLINTNGSHLNTALNCNDSLNTGWLGLFYHHKFNSSGCPERLASCRPLTDF 60
DB 73 GAKNQVLINTNGSHLNTALNCNDSLNTGWLGLFYHHKFNSSGCPERLASCRPLTDF 132
QY 61 DOGWGPISYANGSGDPQRPYCHWHPKPGIIVPAKSVCGPVYCTPTSPVVVGTTRDSCAP 120
DB 133 DOGWGPISYANGSGDPQRPYCHWHPKPGIIVPAKSVCGPVYCTPTSPVVVGTTRDSCAP 192
QY 121 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCGAPPCVIGGAGNNTLHCPTDC 180
DB 193 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCGAPPCVIGGAGNNTLHCPTDC 252
QY 181 FRKHDPATYRCGSGPWITPRCLVDYPRYLWHYPCTINTYTFIKIRMYVGGVEHRLAECN 240
DB 253 FRKHDPATYRCGSGPWITPRCLVDYPRYLWHYPCTINTYTFIKIRMYVGGVEHRLAECN 312
QY 241 WTRGERCDLEDRDR 254
DB 313 WTRGERCDLEDRDR 326

MOLECULE TYPE: protein
US-08-440-103-14

Query Match 100.0%; Score 1494; DB 1; Length 480;
Best Local Similarity 100.0%; Pred. NO. 5e-141;
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAKNQVLINTNGSHLNTALNCNDSLNTGWLGLFYHHKFNSSGCPERLASCRPLTDF 60
DB 37 GAKNQVLINTNGSHLNTALNCNDSLNTGWLGLFYHHKFNSSGCPERLASCRPLTDF 96
QY 61 DOGWGPISYANGSGDPQRPYCHWHPKPGIIVPAKSVCGPVYCTPTSPVVVGTTRDSCAP 120
DB 97 DOGWGPISYANGSGDPQRPYCHWHPKPGIIVPAKSVCGPVYCTPTSPVVVGTTRDSCAP 156
QY 121 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCGAPPCVIGGAGNNTLHCPTDC 180
DB 157 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCGAPPCVIGGAGNNTLHCPTDC 216
QY 181 FRKHDPATYRCGSGPWITPRCLVDYPRYLWHYPCTINTYTFIKIRMYVGGVEHRLAECN 240
DB 217 FRKHDPATYRCGSGPWITPRCLVDYPRYLWHYPCTINTYTFIKIRMYVGGVEHRLAECN 276
QY 241 WTRGERCDLEDRDR 254
DB 277 WTRGERCDLEDRDR 290

RESULT 6
US-08-440-103-14
; Sequence 14, Application US/08440103
; Patent No. 5670152
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,103
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 480 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

RESULT 7
US-08-440-542-14
; Sequence 14, Application US/08440542
; Patent No. 5670153
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,542
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 480 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-440-542-14

Query Match 100.0%; Score 1494; DB 1; Length 480;
Best Local Similarity 100.0%; Pred. No. 5e-141;
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAKQNVLIINTGSHLNTALNCNDSLNTGWLGLFVHHKFNSSGCGPERLASCRPLTDF 60
DB 37 GAKQNVLIINTGSHLNTALNCNDSLNTGWLGLFVHHKFNSSGCGPERLASCRPLTDF 96

QY 61 DQGWGPTSYANGSGPDQRPYCHWYPPKPCGIVPAKSVCGPVYCTPSPVVVGTDRSGAP 120
DB 97 DQGWGPTSYANGSGPDQRPYCHWYPPKPCGIVPAKSVCGPVYCTPSPVVVGTDRSGAP 156

QY 121 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCGAPPCVIGAGNNTLHCPTDC 180
DB 157 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCGAPPCVIGAGNNTLHCPTDC 216

QY 181 FRKHPDATYSRCGSGPWITPRCLVDYPRYLWHYPCTINTYFIKIRMYVGVGVEHRLAAACN 240
DB 217 FRKHPDATYSRCGSGPWITPRCLVDYPRYLWHYPCTINTYFIKIRMYVGVGVEHRLAAACN 276

QY 241 WTRGERCDLEDRDR 254
DB 277 WTRGERCDLEDRDR 290

RESULT 8
US-08-231-368-14
; Sequence 14, Application US/08231368
; Patent No. 5756312
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoactive Polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 480 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-231-368-14

Query Match 100.0%; Score 1494; DB 1; Length 480;
Best Local Similarity 100.0%; Pred. No. 5e-141;
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAKQNVLIINTGSHLNTALNCNDSLNTGWLGLFVHHKFNSSGCGPERLASCRPLTDF 60
DB 37 GAKQNVLIINTGSHLNTALNCNDSLNTGWLGLFVHHKFNSSGCGPERLASCRPLTDF 96

QY 61 DQGWGPTSYANGSGPDQRPYCHWYPPKPCGIVPAKSVCGPVYCTPSPVVVGTDRSGAP 120
DB 97 DQGWGPTSYANGSGPDQRPYCHWYPPKPCGIVPAKSVCGPVYCTPSPVVVGTDRSGAP 156

QY 121 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCGAPPCVIGAGNNTLHCPTDC 180
DB 157 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCGAPPCVIGAGNNTLHCPTDC 216

QY 181 FRKHPDATYSRCGSGPWITPRCLVDYPRYLWHYPCTINTYFIKIRMYVGVGVEHRLAAACN 240
DB 217 FRKHPDATYSRCGSGPWITPRCLVDYPRYLWHYPCTINTYFIKIRMYVGVGVEHRLAAACN 276

QY 241 WTRGERCDLEDRDR 254
DB 277 WTRGERCDLEDRDR 290

Query Match 100.0%; Score 1494; DB 1; Length 480;

Best Local Similarity 100.0%; Pred. No. 5e-141; Mismatches 0; Indels 0; Gaps 0; Matches 254; Conservative 0; 0;

QY 1 GAKQNVQLINTNGSWHLNSTALNCNDSLNTGLAGLFYHHKFNSSGCPERLASCRPLTDF 60
DB 37 GAKQNVQLINTNGSWHLNSTALNCNDSLNTGLAGLFYHHKFNSSGCPERLASCRPLTDF 96
QY 61 DQGWGPISYANGSGDPQRPYCHWHYPPKPGCIVPAKSVCGPVYCFPTSPVVVGTTRDSGAP 120
DB 97 DQGWGPISYANGSGDPQRPYCHWHYPPKPGCIVPAKSVCGPVYCFPTSPVVVGTTRDSGAP 156
QY 121 TYSWGENDTDFVLNNTRPPLGNWFGCTWMNSTGFTKVCGAPPCVIGAGNNTLHCPTDC 180
DB 157 TYSWGENDTDFVLNNTRPPLGNWFGCTWMNSTGFTKVCGAPPCVIGAGNNTLHCPTDC 216
QY 181 FRKHDPATYSRCGSGPWITPRCLVDYPRYLWHYFCTINTYIFKIRMYVGVGVEHRELAACN 240
DB 217 FRKHDPATYSRCGSGPWITPRCLVDYPRYLWHYFCTINTYIFKIRMYVGVGVEHRELAACN 276
QY 241 WTRGERCDLEDRDR 254
DB 277 WTRGERCDLEDRDR 290

RESULT 10
US-08-824-057-3
; Sequence 3, Application US/08824057
; Patent No. 6121020
; GENERAL INFORMATION:
; APPLICANT: SELBY, MARK
; APPLICANT: HOUGHTON, MICHAEL
; TITLE OF INVENTION: NOVEL HEPATITIS C E1 AND E2 TRUNCATED
; TITLE OF INVENTION: POLYPEPTIDES AND METHODS OF OBTAINING THE SAME
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CHIRON CORPORATION
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/824,057
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/282,959
; FILING DATE: 29-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0987.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 663 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-824-057-3

Query Match 100.0%; Score 1494; DB 3; Length 663;
Best Local Similarity 100.0%; Pred. No. 7.6e-141; Mismatches 0; Indels 0; Gaps 0; Matches 254; Conservative 0; 0;

QY 1 GAKQNVQLINTNGSWHLNSTALNCNDSLNTGLAGLFYHHKFNSSGCPERLASCRPLTDF 60
DB 43 GAKQNVQLINTNGSWHLNSTALNCNDSLNTGLAGLFYHHKFNSSGCPERLASCRPLTDF 102
QY 61 DQGWGPISYANGSGDPQRPYCHWHYPPKPGCIVPAKSVCGPVYCFPTSPVVVGTTRDSGAP 120
DB 103 DQGWGPISYANGSGDPQRPYCHWHYPPKPGCIVPAKSVCGPVYCFPTSPVVVGTTRDSGAP 162
QY 121 TYSWGENDTDFVLNNTRPPLGNWFGCTWMNSTGFTKVCGAPPCVIGAGNNTLHCPTDC 180
DB 163 TYSWGENDTDFVLNNTRPPLGNWFGCTWMNSTGFTKVCGAPPCVIGAGNNTLHCPTDC 222
QY 181 FRKHDPATYSRCGSGPWITPRCLVDYPRYLWHYFCTINTYIFKIRMYVGVGVEHRELAACN 240
DB 223 FRKHDPATYSRCGSGPWITPRCLVDYPRYLWHYFCTINTYIFKIRMYVGVGVEHRELAACN 282
QY 241 WTRGERCDLEDRDR 254
DB 283 WTRGERCDLEDRDR 296

RESULT 11
US-08-444-818-89
; Sequence 89, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
; APPLICANT: Chien, David Y.
; APPLICANT: Rutter, William J.
; TITLE OF INVENTION: NANBV Diagnostics and Vaccines
; NUMBER OF SEQUENCES: 777
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,818
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,590
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Ailsa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0110.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508)359-3876
; TELEFAX: (508)359-3885
; INFORMATION FOR SEQ ID NO: 89:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2772 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-444-818-89

Query Match 100.0%; Score 1494; DB 3; Length 2772;
Best Local Similarity 100.0%; Pred. No. 5e-140; Mismatches 0; Indels 0; Gaps 0; Matches 254; Conservative 0; 0;

QY 1 GAKQNVQLINTNGSWHLNSTALNCNDSLNTGLAGLFYHHKFNSSGCPERLASCRPLTDF 60
DB 292 GAKQNVQLINTNGSWHLNSTALNCNDSLNTGLAGLFYHHKFNSSGCPERLASCRPLTDF 351

QY 61 DOGWGPISYANGSGDQRPYCHWHYPPKPCGIVPAKSVCGPVYCFTPSPVVGTTDRSGAP 120
|||||
Db 352 DOGWGPISYANGSGDQRPYCHWHYPPKPCGIVPAKSVCGPVYCFTPSPVVGTTDRSGAP 411
|||||
QY 121 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCAGAPPVIGGAGNNTLHCPTDC 180
|||||
Db 412 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCAGAPPVIGGAGNNTLHCPTDC 471
|||||
QY 181 FRKHDPATYSRCGSGPWITPRLCLVDYPYRLWHYPCTINTYTFKIRMYVGVGVEHRLAECN 240
|||||
Db 472 FRKHDPATYSRCGSGPWITPRLCLVDYPYRLWHYPCTINTYTFKIRMYVGVGVEHRLAECN 531
|||||
QY 241 WTRGERCDLEDRDR 254
|||||
Db 532 WTRGERCDLEDRDR 545
|||||

RESULT 12
US-08-444-818-138
: Sequence 138, Application US/08444818
: Patent No. 6150087
: GENERAL INFORMATION:
: APPLICANT: Chien, David Y.
: APPLICANT: Rutter, William J.
: TITLE OF INVENTION: NANBV Diagnostics and Vaccines
: NUMBER OF SEQUENCES: 777
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Chiron Corporation
: STREET: 4560 Horton Street
: CITY: Emeryville
: STATE: CA
: COUNTRY: USA
: ZIP: 94608-2916
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/444,818
: FILING DATE:
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/403,590
: FILING DATE: 14-MAR-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Harbin, Alisa A.
: REGISTRATION NUMBER: 33,895
: REFERENCE/DOCKET NUMBER: 0110.002
: TELEPHONE: (508)359-3876
: TELEFAX: (508)359-3885
: INFORMATION FOR SEQ ID NO: 138:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2995 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-444-818-138

Query Match 100.0%; Score 1494; DB 3; Length 2995;
Best Local Similarity 100.0%; Pred. No. 5.5e-140;
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAKQNVOLLNTNGSWHLNSTALNCNDSLNTGWLGLFVHHKFNSSGCGPERLASCRPLTDF 60
|||||
Db 406 GAKQNVOLLNTNGSWHLNSTALNCNDSLNTGWLGLFVHHKFNSSGCGPERLASCRPLTDF 465
|||||
QY 61 DOGWGPISYANGSGDQRPYCHWHYPPKPCGIVPAKSVCGPVYCFTPSPVVGTTDRSGAP 120
|||||
Db 466 DOGWGPISYANGSGDQRPYCHWHYPPKPCGIVPAKSVCGPVYCFTPSPVVGTTDRSGAP 525
|||||

QY 121 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCAGAPPVIGGAGNNTLHCPTDC 180
|||||
Db 526 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCAGAPPVIGGAGNNTLHCPTDC 585
|||||
QY 181 FRKHDPATYSRCGSGPWITPRLCLVDYPYRLWHYPCTINTYTFKIRMYVGVGVEHRLAECN 240
|||||
Db 586 FRKHDPATYSRCGSGPWITPRLCLVDYPYRLWHYPCTINTYTFKIRMYVGVGVEHRLAECN 645
|||||
QY 241 WTRGERCDLEDRDR 254
|||||
Db 646 WTRGERCDLEDRDR 659
|||||

RESULT 13
US-08-440-103-36
: Sequence 36, Application US/08440103
: Patent No. 5670152
: GENERAL INFORMATION:
: APPLICANT: Weiner, Amy J.
: APPLICANT: Houghton, Michael
: TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
: NUMBER OF SEQUENCES: 45
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Chiron Corporation
: STREET: 4560 Horton Street
: CITY: Emeryville
: STATE: CA
: COUNTRY: USA
: ZIP: 94608
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/440,103
: FILING DATE: 12-MAY-1995
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/231,368
: FILING DATE:
: APPLICATION NUMBER: US 07/759,575
: FILING DATE: 13-SEP-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: McClung, Barbara G.
: REGISTRATION NUMBER: 33,113
: REFERENCE/DOCKET NUMBER: 0205.001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (510) 601-2708
: TELEFAX: (510) 655-3542
: INFORMATION FOR SEQ ID NO: 36:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3011 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-440-103-36

Query Match 100.0%; Score 1494; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. No. 5.6e-140;
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAKQNVOLLNTNGSWHLNSTALNCNDSLNTGWLGLFVHHKFNSSGCGPERLASCRPLTDF 60
|||||
Db 406 GAKQNVOLLNTNGSWHLNSTALNCNDSLNTGWLGLFVHHKFNSSGCGPERLASCRPLTDF 465
|||||
QY 61 DOGWGPISYANGSGDQRPYCHWHYPPKPCGIVPAKSVCGPVYCFTPSPVVGTTDRSGAP 120
|||||
Db 466 DOGWGPISYANGSGDQRPYCHWHYPPKPCGIVPAKSVCGPVYCFTPSPVVGTTDRSGAP 525
|||||

Tue Mar 6 12:02:34 2001

121	TVSGENDTDVFLNTRPPLGNWFGCTWMNSTGFTKYCGAPPCVIGAGNNTLHCPTDC	180
QY		
526	TVSGENDTDVFLNTRPPLGNWFGCTWMNSTGFTKYCGAPPCVIGAGNNTLHCPTDC	585
Db		
181	FRKHPDATYSRCGSGPWITPRCLVDYPYRLWHYPTINTYTFKTRMYVGVGVEHRLAASN	240
QY		
586	FRKHPDATYSRCGSGPWITPRCLVDYPYRLWHYPTINTYTFKTRMYVGVGVEHRLAASN	645
Db		
241	WTRGERCDLEDRDR	254
QY		
646	WTRGERCDLEDRDR	659
Db		

RESULT 14
 US-08-440-542-36
 ; Sequence 36, Application US/08440542
 ; Patent No. 5670153
 GENERAL INFORMATION:
 APPLICANT: Weiner, Amy J.
 APPLICANT: Houghton, Michael
 TITLE OF INVENTION: Immunoreactive Polypeptide Compositions
 NUMBER OF SEQUENCES: 45
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Chiron Corporation
 STREET: 4560 Horton Street
 CITY: Emeryville
 STATE: CA
 COUNTRY: USA
 ZIP: 94608
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/440,542
 FILING DATE: 12-MAY-1995
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/231,368
 FILING DATE:
 APPLICATION NUMBER: US 07/759,575
 FILING DATE: 13-SEP-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: McClung, Barbara G.
 REGISTRATION NUMBER: 33,113
 REFERENCE/DOCKET NUMBER: 0205.001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (510) 801-2708
 TELEFAX: (510) 655-3542
 INFORMATION FOR SEQ ID NO: 36:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3011 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-440-542-36

```

526  TYSWGENDTDFVLNNTRPPLGNWFGCTWNNSTGFTKVCGAPPCVIGGAGNNTLHCPTDC 585
181  FRKHDPDATYSCRGSGPWITPRCLVDYPYRLWHYPCTINTIFKIRMYCGVHRLEAACN 240
586  FRKHDPDATYSCRGSGPWITPRCLVDYPYRLWHYPCTINTIFKIRMYCGVHRLEAACN 645
241  WTRGERCDLEDLDRDR 254
646  WTRGERCDLEDLDRDR 659

RESULT 15
US-07-910-760-10
; Sequence 10, Application US/07910760
; Patent No. 5683864
; GENERAL INFORMATION:
; APPLICANT: Houghton, Michael
; APPLICANT: Choo, Qui-Lim
; APPLICANT: Kuo, George
; TITLE OF INVENTION: Combinations of Hepatitis C virus (HCV)
; TITLE OF INVENTION: Antigens for Use in Immunoassays for Anti-HCV Antibodies
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: P.O. Box 8097 (Int. Prop. R-440)
; CITY: Emeryville
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 94662-8097
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0; Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/910,760
; FILING DATE: 07-JUL-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Blackburn Esq., Robert P.
; REGISTRATION NUMBER: 30,447
; REFERENCE/DOCKET NUMBER: 0101.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2702
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3011 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-910-760-10

```

Tue Mar 6 12:02:34 2001

us-09-407-430-3.ra1

Page 9

Db 586 FRKHDPATYSRGSGPWITPRCLVDYDYLWHPCTINYTIKIRMYVVGVEHREACN 645

Qy 241 WTRGRCGLDRDR 254

Db 646 WTRGRCGLDRDR 659

Search completed: March 6, 2001, 11:52:27
Job time: 105 sec

